

Fig. 1a

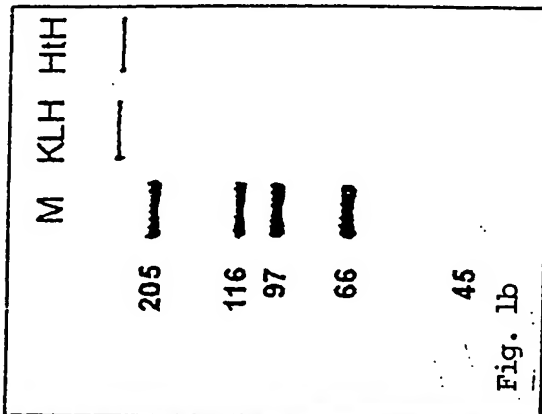


Fig. 1b

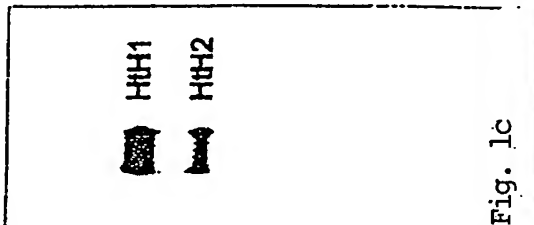


Fig. 1c

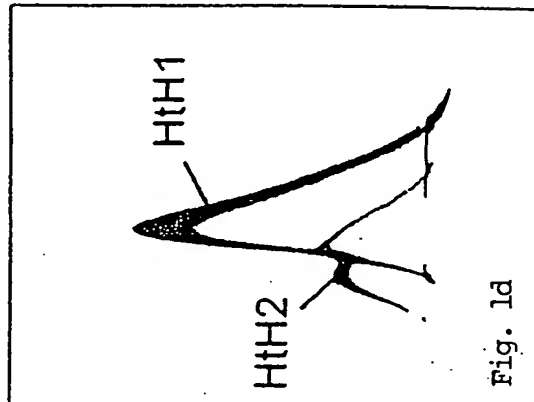


Fig. 1d



Fig. 1e

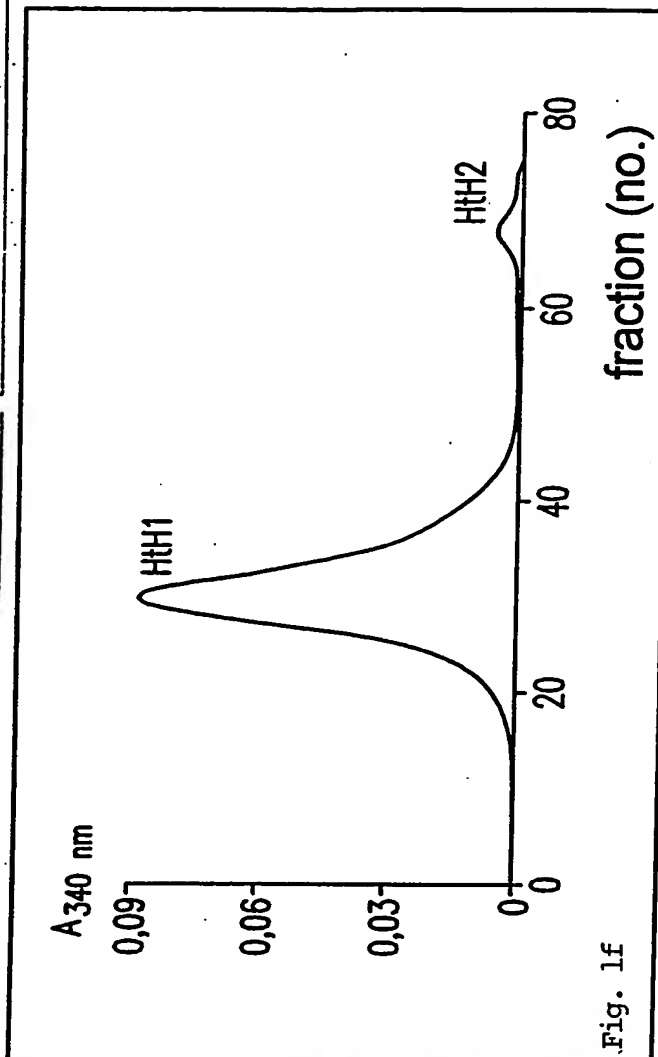
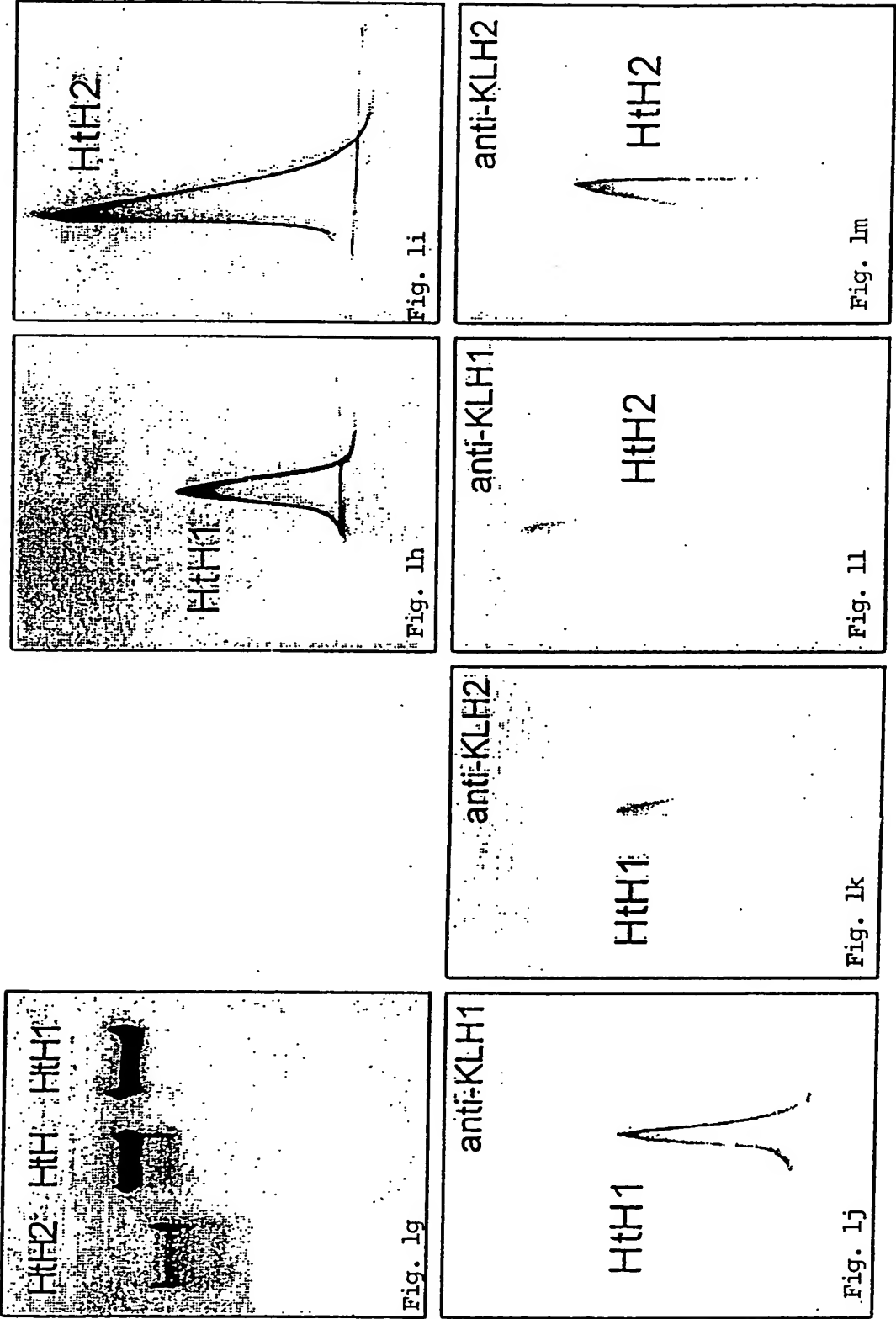
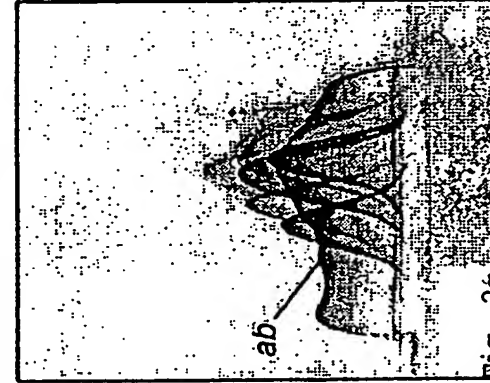
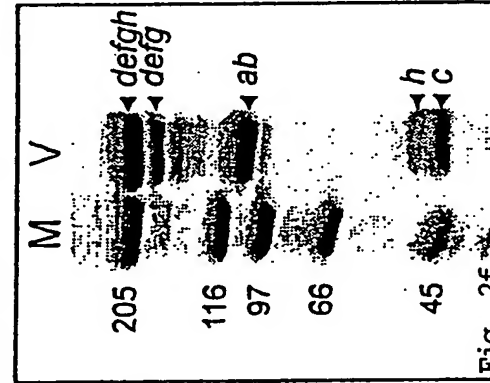
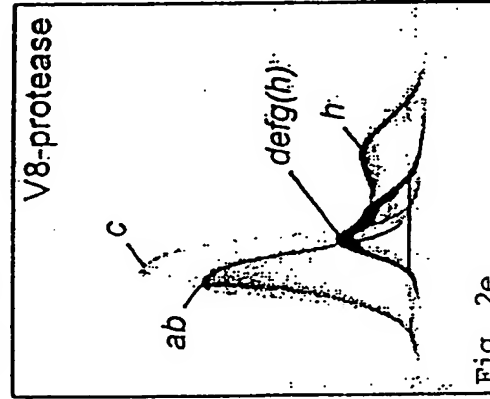
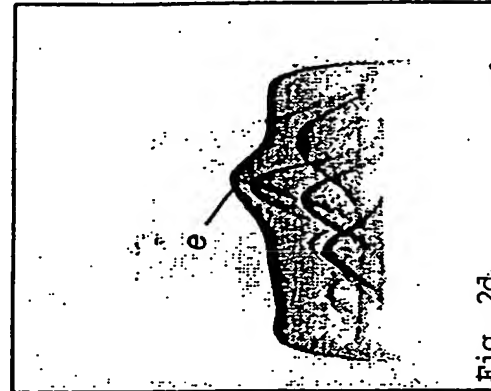
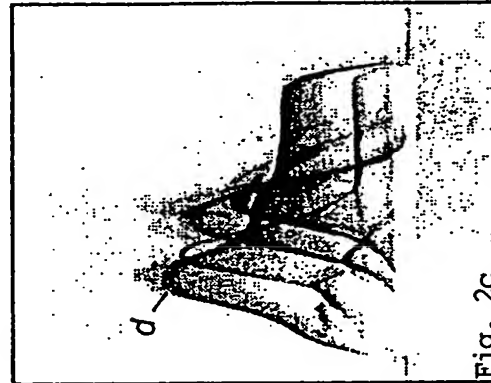
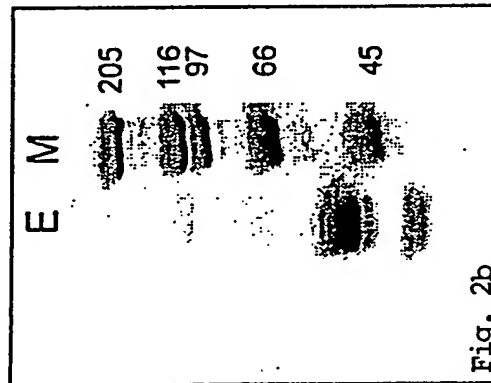
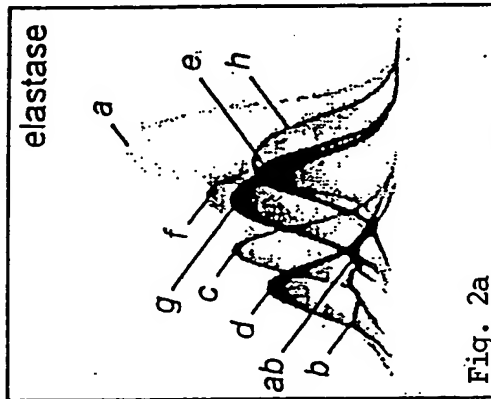


Fig. 1f

Nucleic Acid Molecule Comprising A Nucleic Acid Sequence Which Codes For A Haemocyanin And Comprising At Least One Intron Sequence (Serial No. 10/049,988; Inventor: Markl Jurgen)





Nucleic Acid Molecule Comprising A Nucleic Acid Sequence Which Codes For A Haemocyanin And Comprising At Least One Intron Sequence (Serial No. 10/049,988; Inventor: Markl Jurgen)

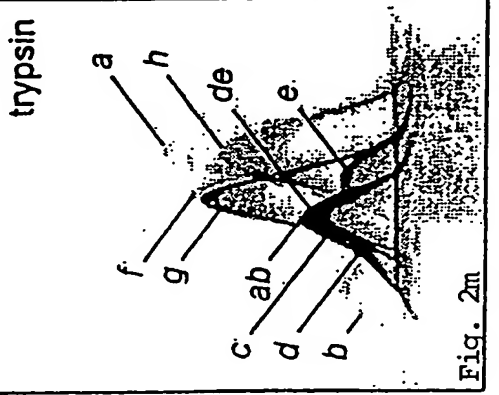
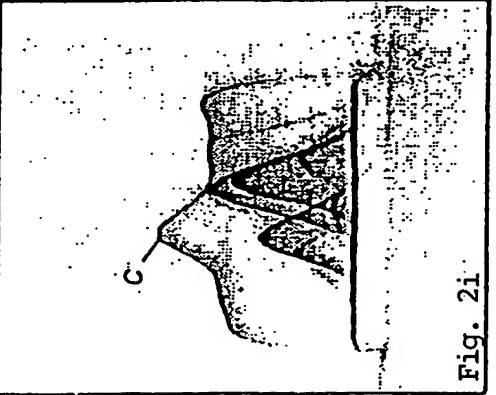
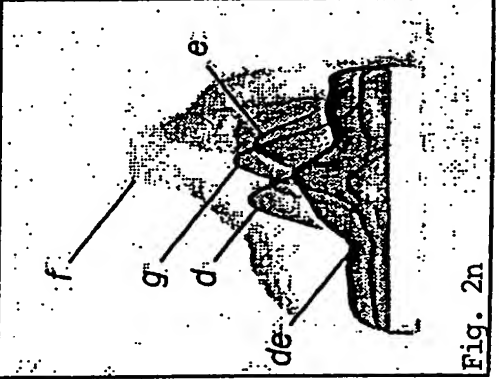
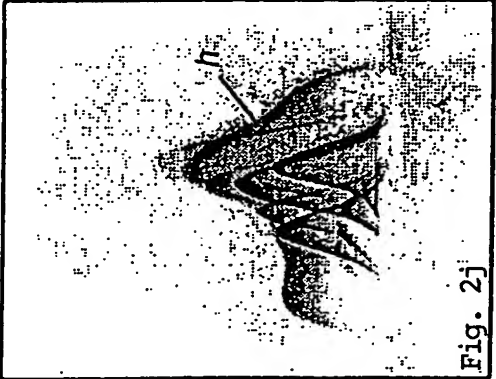
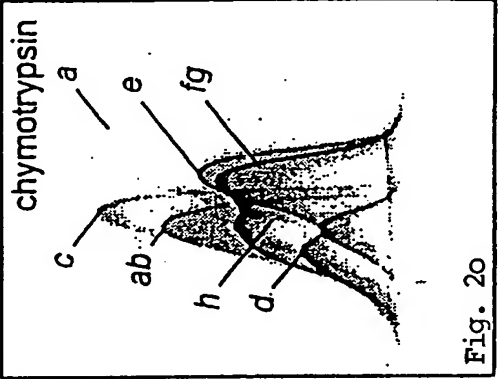
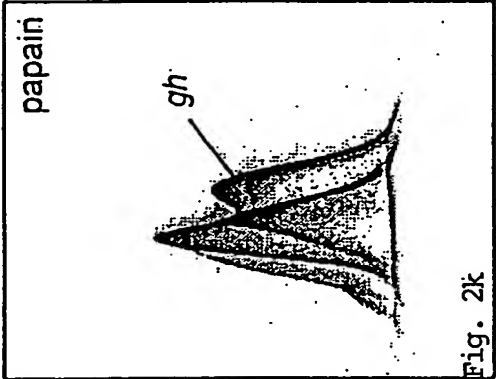
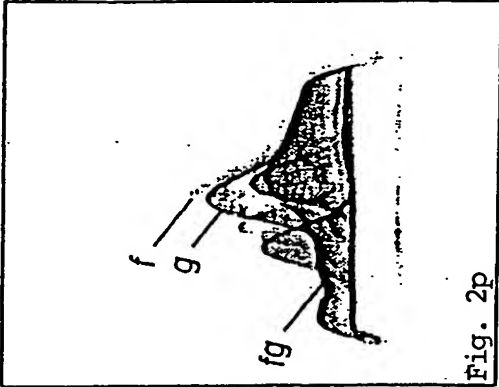
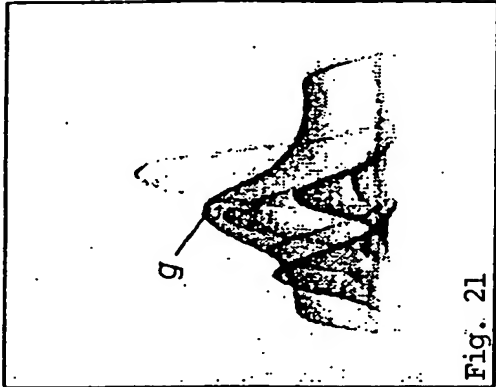


Fig. 3a

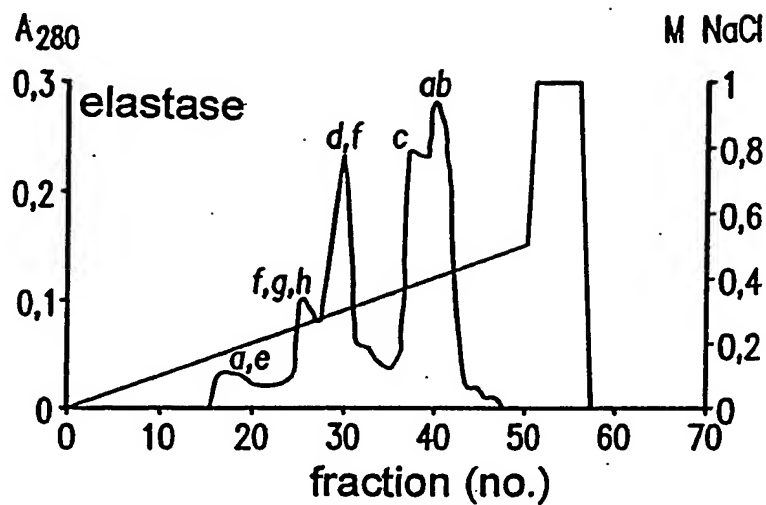


Fig. 3b

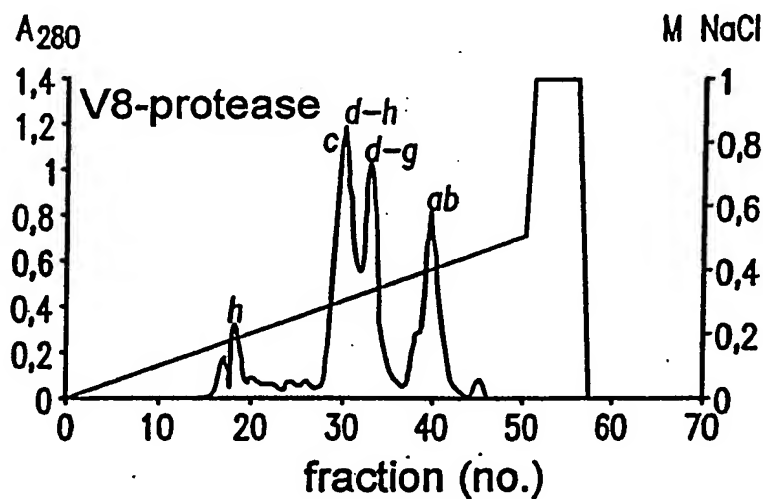


Fig. 3c

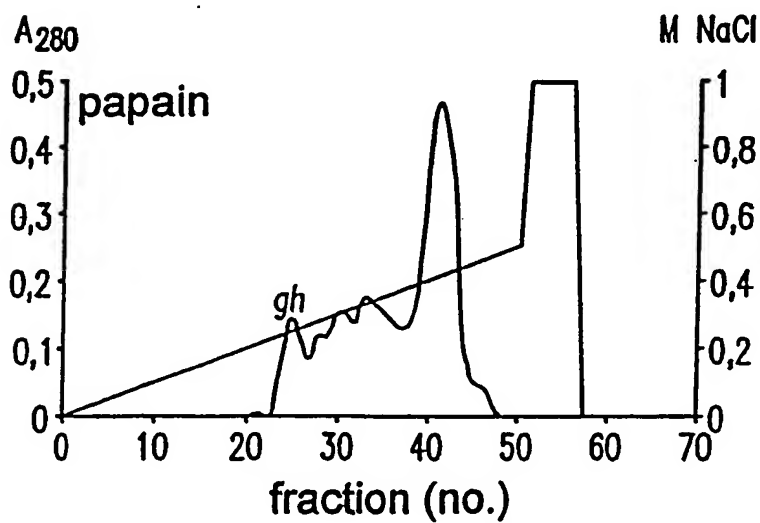


Fig. 3d

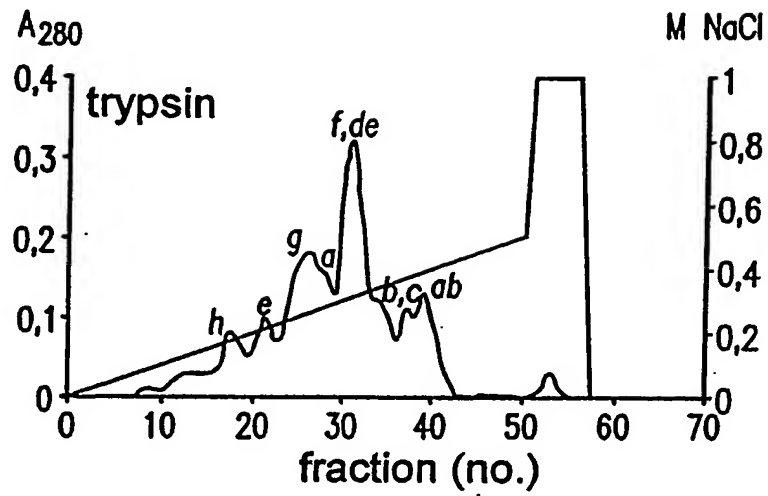


Fig. 3e

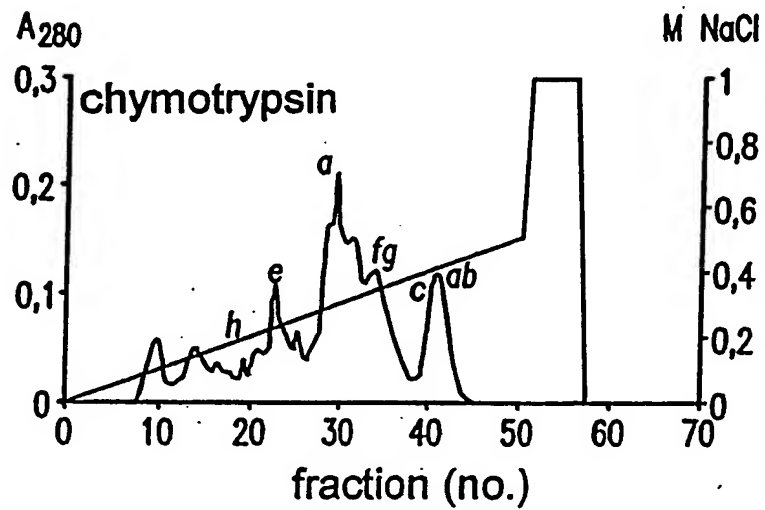


Fig. 4aGenomic sequence of the HtH1 gene

SIGNAL PEPTIDE SEQUENCE 1S-1 (1st part)

GGCTTGTTTCAGTTTCTACTCGTCGCCCTTGTG

INTRON 1S-1/1S-2 (SEQ ID NO:109)

GTAAGTCAACGTCTTTGTTTTAAGTTTGATGCATATCTATCATTGCGTTTTAAAATACCA
TTACAACCAACGTGTCTCTATTGGTCTTCACCTGTTTAACGTATATATTGTTTTTAATGT
GAAAATCTGAGATTATTTTCATTTCCGTCAATATTCGTAAAATACTATACAAATAAAATT
GCTTCAGCCTATTGCATTGGCAGTTTTTCGCAGAATAACGAGGGAAGGCGTACATAAAATA
TAAACCAAGTGTATATTCAAGCATGTTTATAATTTCTTTATAGATTATAACATCATATCAA
AACACCAATCTGGATTTAAACCCGTGAATCCAAAGTATACCAATTAACGGAACCTTTATCA
TGTTTTTATCAAAGGTTTTAGATGAGGGTAAAGAAGTCCGAGCTATATTTTGCGATATCAG
CAAAGCCTTCTATCACGTCTTGCACACAGGGCTGGTATCTAAACTCGAATCCACAGGAAT
AAATATTTTCAGCCGATAGAGAACAGTCGGTGGCTATCATTGGTCACAAAACAAGTCCAAA
ATCTGCATTAGCCGGTGTTCCTCAAGGCTCTGTCTTGGGGCCACTATTATTTCTCACCTA
TATAAACGATTCAACTAATGGAATATAAAGCAACGTAAACCTCACCGCAGATGAAACACT
AAGTTATAGACAATCCGTTTTAAAACCCAGCCACTGCTTAATAATGACTTAGGCCGTCTTT
CAGACTGGGCTAGTAAGCGGCAGGTTAAATTTACCTTGAAAAGACAGAAACCATGGTAT
ATTTCAAAAACACGAATGCAAGTCCTAAACTTCAACTACTACTTGTATGATACTGGGATTT
CTAAAGTGTGTGAACAAAAACACATTGGCCTGATCCTACAAGATAACCAGACAGAAACCA
TGTTTTTTTTCAATAACACGAATGCAAGTCCTAAACTTCAACTACTACTTGTATGATACTG
GGATTTCTAAAGTGTGTGGTGAACACAAACACCTTGGCCTGATTCTGCAAGATAATGGAA
AATGTCAGAAACATAAGCAAGTTGATGTGGGGTTTTCTGGGGGTGTGACAACACCGAAA
GACCCTGCAACTAATGTTAGCTCAAAGGGTTTTACACCCGGTCACAAGTGGGGATCGACC
CAGGCACCTTTTGCCTTTGACAGCTCGCCTTTCAAAAAATCTCAATTCGAAAACGAAATC
TAATAATTTTCATGAGCGATACAACCGTTTTTTCATAATGCTGTGGTACCGCATACTGTGGA
AACATCTGTCTACCCATTTGGTAGTCCCCCATAAAATGTATTTATGTTTATAAACACAAT
GTTTATAGGGTTACAGTTAGAAGAAGCATTTCTATTGGCTAATGTACATTGCTTGTTTTTT
ACTATTGTGCAAAGGCATATTACAGGTCTTTTAGGAAATTAATACTGTTTAAATCACAT
ACACTACCGGTAATCCTATTATGCTTATCCTGCCAACATTCTGCCCAAGCAAACGCATGA
AAGTTAAAGCTGAGTGTAATAACTGATTGCTGTGTTACTTCACAACCAGTGGACTGAAT
ACAACCATGTTTTTTCTTGAAAGTCACAAACATCCAGTCGGTTTCTAATGTGTAAAGTTT
CTAGTTTCATAAAGAGCATGACGTAATGGTGAATAGGAGTTATCAATGTTTCTATCTAAT
GACTCCTAGTTCGTTACTTTTTTAATAAAACATCCATGTGTTTAATGTTTGGCCACAGAT
ATAACAAGAAAGAAATCGGATAAAATCTACATTTTGACCAATCGGAAGGCTGCCCCCTCC
CTAATCCTAATCATTTTTGTGCCTCAAAACATACTCAACCAGACATTTGAACTATGTATA
TATCAGAATGAAATGGTAACAATAAACTTGTATGTTGACCAGACAGAATTAGGGTGAATC
TGAATACCAACTATTGTACATATGAATATGGATAAGCTCTGCGCGTGCGTGCGGGCGGT
GTAGTGCGTGTGTGTCTGT
TGT
TGCACAGACATGTGGTTGAGACACACTTGATTTCAGTGCAGGATTATGTCCTTCAACCGAG
TGTAAGTCTTTAAGTGTGCCTGGAAACAAAAAAGTGCCTTGGGTTGCATCGCCTCTGTAGC
AAGCTTGGACGCGTCACGCAGCTCTGATACCACGTATTGGCACCATTGTTTCATCGGTCTC
ACGCGAATATTATGCTATGTGTGGCGTATCATAACCATAGGTTGGGAACGTTTCAATACTG
TACCGAGCTTGGGCGTGTCACAAAGCTATGATAAGATGACAACACGTCTTGGCATCTTGT
TTCTCGGTATCACGCGCTGTTATGCTATGTGTGGCTATCACACCTTAGGTTGGGAAAGT

Fig. 4b

TTCCACATTTTCCAGCCTCGTACATGTTTCCTTTTGTTTTTTCCTTAGTTATCAGCATAC
CGTATATTCTATATTTAATGAGCATTGTATTTTTCTACAG

SIGNAL PEPTIDE SEQUENCE 1S-2 (2nd part)

GTGGGGGCTGGAGCAG

INTRON 1S-2/1A-1 (SEQ ID NO:110)

GTGAGTTTCTTAACATTGTCATGGTACATGGATATACGCTCAGTGGGAAAGCAGGATATC
CCCTTGGTTCAAGTATTCACTTGTACGCCAAGTGTTTCGATTCCCAACATGGAATACTGT
CATATAGTAAATTGATACACTACTTACATTTAATTCTCCACTAAACGTCAACGTCTTTTA
CTTCATGGCCACATGGTCCGTATTAGTGAGTGAGTGAGTCAGGGCATAAGTATTTAACG
TCAAATCAGCAATATTTTCAGCCATATTGTGACAAGAATTGAATATAAAATAATTATACTTA
TAATGCTTATAAATATAAATTATATAAATACCTATAACTATAAATTAGTTATACTAGTAT
TTATCAAAACATATTTGCCACGACACTGCACGCCGATACTTCAAGTGTCTTCACCTCAAG
CGTGTAACCTCCTCATACTCTGTAATAAGTATGTACACTAAGTGAGTGCTATCATCTCCAT
GCTTCATTAGTTTTCGTCAGATGCGTGTATCCATACGAGTACATTTCAGATTATGGGATCCA
GAGCTTTCTTATCTCAAGTATTTCCGATTGTAAAGCCATACTACTTCCCCAATGACTGAC
GAGACAGATGGCAACCGTTCTTTCTCCTGACTAGGTGAGTGCCACTGATAAATCATTAT
GCCTTTAACATTAGGAATGTTAGCAGTGACATGTTTCAGAATTGCGACCTTATGGTTGT
AAAGATTACAACTTTTACAACCTTACTTGAGACAGGTTCCATATGTTCGTATCTGAAATAGT
GTGAAGGTATCTGATTTCGATGCAATACACAGACATATAAACATATTGTTCGCCCTGCTATT
CCGGAAAGGTCATTTTGTATGTAACGTTCCCTAATGGACACAAACCGGAATTATTAGTTAA
ACATACTCAACAAAACCTATGTTATTTTGCAATGGGTAGCACCGAAATCTACCGACAGTGG
TTCGTAAAGTAGAACATTCTGACATAAAGAAAAATCATTGGCTTTAAATATATGCAAGT
TACTTGTCTCTAACAACCAGTTTTTATACACATTTTCAGAGAACGGGGAATCCGCGATGACA
ATATCAACGAGTATATACAGAATATATAATTAAAAACGATGAGTGCCTGGCAAGGGAAAG
AGCGAGATTTGCCAAACAGGGGGGTGGTGGTTGAGCTTGAATCGTGGAGAAACGTAGATTG
AAAGACAAGATGACATCTAATGATCCGAAAATCAAACACAGGATTAACCTGGGATGCAGAA
GAATGAATATCTCAAGCATAACATGCAACACTTCATGAATGCATCTCAAACATTTTTCGTCA
GATCGGATGCATGAAGATTTGTAAAGCAATGGTTTAAATTGTCCCTAAACGTTTAGTTGG
AGATGTATGAGGCTAGGCTGTATGTTGAACGAAACCATTTAACATTGTTGTTTCATGATTA
TTTAATATTTTTTTCATTTTATAGATGTACAATAAAATTGGAACTAAACATTTCCCTTTA
TTGTTTTGTATTTACCTGTTTCATGGGTATGTTTTGAAAGATCGTGATATTTAGTTGGCAT
TCACAAGTTGGAAAAAGGTCACTCAGTTTTCAGTTTATGTAACCTCTTTATCTGA
CGCTCCAAAATATGTATAGCCTTGTTTCATCTGTTCGGTATGTGGATATTCCTACTTCAGGG
TAGGGTAGCATTAATACTTACAAAACATAACGTGTACCAGATTTTCAGTCACCTCAGAGAT
GATAATGCATGTTCGATATGATAGGTCAAACCTTTTCGATATCAATCACAATGAACCTATGG
ACCTGAAATCGGAATGATACGTTACACTTTAGAAACAATTCACAAATATGACTGTCACCC
TTTCAGGTAATAATGTTTGACGGACTACGATAGTGCTGAACAGCAGGAGAGGCAACATGG
TTCGATTGTGAGACAGGTTTAGTGTATTTGTTTGCGAATTTAAGGTTCTGAATCACAATA
GACACGGTTCAGTTAATGGATAAACCAATCATTAGATAGATAGAGATTAGTCGCGATATT
GCTGGGATAAAGCTTAGTGGGACGTTAAGTCCCATCTCAATCTCTCATTTTTTCCAAA
ACAGTTTTTAATTCAGGCTCATGACAAGGTCGTAAGTTGCAAAGGATTCTACTTCAAGCA
GAGATGTCTCATGAATACAGTACAGGTTTTTGAAGTTTATCCAGTGCAGCGCTGGCACC
ATCTCTGCATGCGAATTATACCATCCATGCCGCTCTAGGCTATTTGTATTAAGTCTGTAG
AATTAAATTCGCGAGTTGCAAATACTGCTCACCATTATCTGCCTCAACCCAGTTTGGGTA
CATCCAAAGTTTTGGGAGATTTTATTCGAGAAATCAACCTGAGATGTTGAATCGGGAGCT
GCGCTTATTCAATGGTGGACTCGGAAGGGAAGTAACCGCTGATGAGGCAAAACAATAACG

Fig. 4c

CAACATATGGAAGTGGAACCTCTTTGAACCAGTATTATGTTTGTGTGGACATGTATGTGT
TAATTTGACCATTTCGAACAACCTTTACTATTCTATTCCATAATGTGTTTAGATTTACATTTG
AATTAAAAGAGATGAGTTTAAAGATATTAATATTTTCCTTTTATAGTCTGTCTGATTGTA
GGGCAATATTTATGTATGTTTCGTTTCATTTTTCATTTATCATTGGAAGGTATATCATAA
GATTATTATTATCATTCTTGAAGTAATGTATACATATATATATGTCTTGAGTAGCTTATT
TTCAATTTATTATCATCCGTCATCCAATTTTATTTTACGAAAGTATAAGAAATAACGAGA
GAGAGAGAGAGAGAGAGAGAGAGAAAAGACAGAAATGAAGTTAGGAGATATNAGTTATCAAGAA
AACAACAGTTTGAATTTTTTGTGTTAGACAAGATATCATATCAATAACCTCGCACTATTAC
GGGAATAGGCGGGCGTTCCATATGCACAATGAATCGTCAGTTAAAATCAACATTAAACTT
AAAATACTCCTCATATTTAAAGTTGATCTACCTCTTGTATTATTGTAGACTATTAGACAG
AAGTCGACAGTGACACCAGCAACCAGATATCATACCCAGACTTAAAAAGCTGTTTCCTTG
ATGTTTCAATTTATTTCCATTTCCATTATTTCCCTTTTATTGGTTTCCATTTATCAAACCTT
ACCATCTGCACCAGTGGAAGATTGATATGTTGTATTTATTTATATTTCTTGTACTACAAT
ATCAAGAATGTATAGGAGCTATTCCCTTGTTCCTAAAACCGGATAGATCCATAATTTCCAT
TTTGGGATAAATGGAACTAAACACAACCTTTTACAGTAAACACGAGTGAGCAAGTTGAGT
TTTACGCCGTTTTTTAGTAGTATTCCAGCAATATCGCGGCGGGGGACACCAGAAATGGGCT
TCACACAGTGAATGCATGTGGGGATTTCGAACCCGGGTCTTCGGCGTGACGAGTGAACGCT
TTAGCCACTAGGCTACCCACCGCCTATTTATAGTTAAGACGAATACTTTTCTCAAGCCT
CAAATATGTCCATTCTAGAGAGACTGAATCTGATCCTGAATCTGCGGACCGGTCTTGAAT
ATCATCCCACTAACTCATTGTACAAAGTACCTGTAGATTGTGAGTTCAAAGACAGATTTT
ACAACCTTATTATATTTTGTCTGCTCATTAAGATATTCAGACTCACTCAAACCTGCTAAA
TGATTTTAACTCTACTTTGAGATGTTTTAACTTTTATTTCGATGCATTTTTTGCGTTCTGCG
TCCTGTATAAAGGTAAAGCAGGTAAACTAACCTAACCTGTTGATTTATTTTCATAGTTTTTG
CGATCAGATTGAAACCGGAATGCACAGTGAAGTGTGGCATAACATCTTTCCACAGAGATAC
TGGATACTAGGTGGTACAACCGCATTGGCTTTGTGAAAGGATATTAGTGTTTTATGAGAC
TGACTCATGTTTCAATGCTTAGAGCGGAATGATCTCGGTCTTCATGAAAAATATTGTGTT
GAAGTAACCCCCCAGTCCCTAACAGAACGTGGGGAAAGCAGATGGATATGCCAAGACATC
TTCGCATGGTGTGAAGATGATCGTTACAACATCTGCAGAAAAAGTTATTTCTGTGAAGAA
TATGCCAAAGCATCACTGTGAGTGTGTTTGAAGATGTGATATGGCAACACGCAGCGTGTA
TTATGCTTTGTGTGTATTTCTGAAGATCCGTATGAGCATGGCGCCAACTATCAGTTAAA
TGGCTATGCGAAGATCTTCCCGAGATGGTAAACACATATTTTGGCCATTTTCTTTGTAAG
TGGGCGACACAGAAGATCCCCCTGATTGTGTGGATGAGGACACAAAAACGGGTCCCCCTT
CCTTTGCTGATGCTAATGACGCCCTGGAAACATTGAAAGACTTCTTCTCCAGCAAGCAAG
CCACCAACCACAAGTTGTATAAATCGCTTGCGGACTTGAATACGGCAGTTGGACAGATAC
ATACAGCCAGAGAGGGCCGAACATAAACATCTAAACATGGAAAACTGTAAAGACAGGCT
TTGTTGTACGACGTACGTAAATTCATTGAATGTTTGAAGGTTAGAAAATTATTAATCT
TTGAAACCTCGCTCTGTTTGTGTTTGTATTGTCCCCCACATTTGCAAATGGTATCCAAAA
GGGCAGACACATTTGTTTTAATCTTAGCCAGGTTCAATTTAGCCTTGCGCCAGACTCAT
TGTATCTGGTGAAGGCTATAGGTGGCCACGTCTTCTAAGATGCTATGCTATTCTTACCAG
AATCCAATGTAAAGAGTTCAAACGCATGGTTCGCTTTGATTGTGATTCTTTCTTAGCACC
TCTCTCTACCCAGAGTTACCTGCACTGCTCCTGACTCACAATAAGCTGACGTGCTGTC
ATATATGTGCAACATTGTATACGTTGGCGTTAAGCCCAACTCACTTCCGCTGTCTTTTGG
CAG

DOMAIN 1A-1 (1st part of domain a)

ACAACGTCGTCAGAAAGGACGTGAGTCACCTCACAGTTGACGAGGTGCAAGCTCTTCACG
GCGCCCTCCATGACGTCACTGCATCTACAGGGCCTCTGAGTTTCGAAGACATAACATCTT
ACCATGCCGCACCAGCGTCGTGTGACTACAAGGGACGGAAGATCGCCTGCTGTGTCCACG
GTATGCCCAGTTTCCCTTCTGGCACAGGGCATATGTCGTCCAAGCCGAGCGGGCACTGT

Fig. 4d

TGTCCAAACGGAAGACTGTCGGAATGCCTTACTGGGACTGGACGCAAACGCTGACTCACT
TACCATCTCTTGTGACTGAACCCATCTACATTGACAGTAAAGGTGGAAAG

INTRON 1A-1/1A-2 (SEQ ID NO:111)

GTA ACTACAAACGTCGTC CCCATTTCATACAGGAGAAATATACAATTGTGTTGTAAGAGCGG
TATACTGTTTGCCA ACTGTGTAATTGAAACGTTGATGATGGTGTCTTTGTATTTCAATTT
GTATGCACTTAGACATGATCAATGTTTCTGATGTGTCAAGGATGTTTCGGTGTGTCACTTT
CAAAAGATCAAATTCATATGACGTACACAGAGCAAGAACCAACAGTAAGAAGTCTGTATG
ACTTCGCTCTTAAAAGCAATGGAAAAATATTTTCACTTAACACCTAGCCCATATCACGC
ATATTAGATTATTCAAGCGATGTCAACATGTTTTTAATATCAATCTCATGGTTCTGATAT
TACCGGAGACATGCAACAGGCTGCCATTATAGCCAGGAAATCTTATGAATATGTGCATAT
TTTTTCTTTGATTCTGTATGACGAGAAATATTCGGAGGCAAAGATTGTGTTTTTCAGAAC
GAATCAGGGTATCAGTGACATCGTCACTGCATGGCTACAATATTGCTGATGTGACTGTTT
CTCCAAGGATTTTCATCTCACTGTCTGTACTTTGAATCTACAAATTCGTATTAAAGTTAT
GACAATTTTACCCCTGCCTATTTGTAAACGAAATATAACATGAGTGTTTATGCTGACAG

DOMAIN 1A-2 (2nd part of domain a)

GCTCAAACCAACTACTGGTACCGCGGCGAGATAGCGTTCATCAATAAGAAGACTGCGCGA
GCTGTAGATGATCGCCTATTTCGAGAAGGTGGAGCCTGGTCACTACACACATCTTATGGAG
ACTGTCCTCGACGCTCTCGAACAGGACGAATTCGTAAATTTGAAATCCAGTTCGAGTTG
GCTCATAATGCTATCCATTACTTGGTTGGCGGTAAATTTGA

INTRON 1A-2/1A-3 (SEQ ID NO:112)

GTAAGTTTGGTTTACAGTTTTCATTATAAAAACATAGCAGTTTAAAGTTTAGGGGCAGATT
CTAATCTCTAATATTCCTTTCAACTCACTTTATTGGTGCCTTCTTGGAGTGACATTTAGA
AACTAAGACAAGAGGAAGATGAACAATGTTTGTAGGGATAGACAGCTTGGATGCAATTTTC
GGACCAGATTCTAACAGCGTCATGAAGCAAGTGATACACAACGTTATCAATAACGAGAAT
ATACACATAGATGGTTTGAGTTTATAAATGAACTATTAACGGCATTGTGGTTATAGACAG
TGAGGAAGACGCCAGATAGACAAAGGGTAGGGGCCCTTGGTTAGATAATGAGAAGTTGAAG
AGGTGTAATAACTTAAATCTCTCTTGACTATTGATTGTGTCTAAGAGTTTTCTTATCTTA
CAGTCGGCCAGTTGGGTCAAAGATGGTGTGATTCCGATGTGCTTTGTGTGTTCTGCGATG
GCTGATTTAGAGTCAGTTTACTTCAGATGAATGAAGTTCCCCGATTCTTATGTTTAAGTT
TGTTTCACCTACGCATGAAGACATCACCAGCAGGGTCGTCTTTATTTCTAGTAGCTTATT
TACAGCAAGCTTGTAACGTATGCTGAATTGCTGTGCCTCTGTAGAACACAGCATCTATGT
TTGCTTGCTTCTTTAGTAGACTGCGGATGTGATGGTTGGTTACCTGGTATGCTGACGAAA
GAATTGTTGACGTGGTGGTTTGCCTTGATGGGTTTCGTTGACTTGGTTTGTGATACTGA
TTAAGGTGACTCTGCTGGGAGGCTTGGATTCTGGGGCCGGTGTCTTTGCTCTCCTGTCT
AGGGTGGCGATTATTTCCCAACCCACTTGTTCCATTACACTCAAACCTGCTATCAATTT
ACAG

DOMAIN 1A-3 (3rd part of domain a)

ATATTCAATGTCAAACCTTGAATACACCTCCTACGACCCCATCTTCTTCCTCCACCACTC
CAACGTTGACCGCCTCTTCGCCATCTGGCAGCGTCTTCAGGAACTGCGAGGAAAGAATCC
CAATGCAATGGACTGTGCACATGAACTCGCTCACCAGCAACTCCAACCCTTCAACAGGGA
CAGCAATCCAGTCCAGCTCACAAAGGACCACTCGACACCTGCTGACCTCTTTGATTACAA
ACAACTTGATACAG

Fig. 4e

INTRON 1A-3/1A-4 (SEQ ID NO:113)

GTGAGACATTATTACACTTCTATTTAGTAGTGGGGGCGGGATAGCTCAGGTGGTAGAGCG
TCGGCCTTCAGCTTCTAGTCTCGCCACAAGAGCGCGCTGGCTAAAGGGCCGGAGTTAGA
TTCCCGCGGGGCGGCAGGCAATATCTCCGAAGGGGAGAACAGTTCTCCAGTCGGTGAAATT
GGGGTGCAATGTTGTACCACTGAAATGCGTGACACCAACCATCCAAATACCAGCCTTG
CCGCGCTGGTCTGACTACATAGTACCACCCGGATTCAACCGGGCTATATAGGTTCTCCTC
CAGCAGTAAATCTGACAGTCGCCATATAGCTGGGATATTGCTGAGTGCGACGTTAAGCCC
CAACTCACTCACTTTATATTTAGTATTCTATTTAGTATCGACGCATGACCATGTGTGGTG
GTCTACTCATCTCAACACGACCGATTAAACGTTAAGAGCTGCCAACATGATTCTCTTTCTC
TCTTTAGCCTCTTTATGCCAAAAGCTATATATTAATGTAGGACCCTACATATATTATTTTC
CAG

DOMAIN 1A-4 (4th part of domain a)

CTACGACAGCTTAAACCTGAATGGAATGACGCCAGAACAGCTGAAAACAGAACTAGACGA
ACGCCACTCCAAAGAACGTGCGTTTGCAAGCTTCCGACTCAGTGGCTTTGGGGGTTCTGC
CAACGTTGTTGTCTATGCATGTGTCCCTGATGATGATCCACGCAGTGATGACTACTGCGA
GAAAGCAGGCGACTTCTTCATTCTTGCGGGTCAAAGCGAAATGCCGTGGAGATTCTACAG
ACCCTTCTTCTATGATGTAAGTGAAGCGGTACATCACCTTGGAGTCCCGCTAAGTGGCCA
CTACTATGTGAAAACAGAACTCTTCAGCGTGAATGGCACAGCACTTTCACCTGATCTTCT
TCCTCAACCAACTGTTGCCTACCGACCTGGGAAAGGTCACCTTGACC

INTRON 1A-4/1B (SEQ ID NO:114)

GTAAGTTGATTGTCTTAATATTGTTTTAATTTTTGCAGAAATTTGATTTTAAATTGTGTA
ATAACAGTACACATTTTTACGCAACAGCAGTCATTATTGTGTGTGAAGATGTCAAACCAG
AAAGGTTTCAATCGTGAAAACAAAAACAATTCTCTATCTGTATACCCCTCAATACCAGTA
TGATCACAAATCTAGGAAATATTACAATACTGCTTCATAGAGTAACTGCTGTTTGTGGCA
GAGCTGGATACGAAGTTTCTGATAGTTCACAGCTACATGATAGTAAATGAACCTGTACAC
ATCAACGGTTGATCATGAAAATTTTGTATGTGTGAAAGTGCTACCTGTATTAGTGAACGT
GCTACCTGTATAACTGAAAGTGCTACCTGTATGACTGAAAGTGCTACCTGTATGCTGAAA
GTGCTACCTGTATTAGTGAACGTGCTACCTGTATAACTGAAAGTGCTACCTGTATGACTG
AAAGTGCTACCTGTATTAGTGAAGTGCTACCTGTATGAGTGAACGTGCTACCTGTATAA
CTGAAAGTGCTACCTGTATGACTGAAAGTGCTACCTGTATTAGTGAAGTGCTGCCTGTA
TTAGTGAAGTGCTACCTGTATGACTGAGCGTGTTACCTGTATGACTGAACGTGCTACCT
GTATTAGTGAAGTGTAATCTGTATGAGTGAAGTGCTACCTGTATTAGTGAAGTGCTA
CTTGTATTAGTGAAGTGCTACATGTATGACTGAAAGTGCTACATGTATGAATGAGAGTG
CTACCTGTGTGACTGAAAGTGCTACCTGTATTAGTGAAGTGCTACCTGTATGACTGAAC
GTGCTACCTGTATTAGTGAAGTGCTACCTGTATGACTGAAAGTGCTACCTGTATGACTGA
GAATATCTGGGCTCAAACAGTTTTTTCAGTATCATAGTCGTATCAGTTTGATTTGTATGT
GCAGTGGAATCATTTTTCGTCAAATAATCAAACCTGGTGTTGAACTGGCGTTACGTTTTTA
TGGTTGTAAAACAAATTCTGTAAGTAAAGATATTTTAGGGATATCTGTATGACATGAACT
GAATTGCTTAAGGTTAGCATGCCATGACAAATTGCTGAATGTCTGAGGATTGGTGGAGCA
ATAAATCATTATTAAGACAAAATCAGAAACGTCCATTTTCACTTTTAACAGTGTATCTG
TCTGAATGCCCCCTACTTTTTGGAAGAGTATATATGAATTATCGGCAATATAAACGTTA
AATGGCAAATGTGCGGCATATGTCAGGACATTATTACCGCAGTTTATAGTCATATTTACC
GGGTCTAGGACAATTGTCACCCCGACAATTGCCACCCGGACAATTGCCACCCAAAAATAA
AATATACGTAAACAGAAAACAAATATTGCTTTCAGCCTTTATTGAGTTAGATAATGACAT
TTATGTTGATAAATATGTCGTTTGATAATAATAACAATAATATAATATTACAATACT
GCAATAGTACTATCAGTACTTATCATTTTTATCACAGATTATATATAGATTCTAGAGTCCG

Fig. 4f

ATGTTGTAGGCAACACTTCGTCGGTAGGCCGTTAGGTAGTTATCATTAGGGCTGAGTATT
GCGCCAAATTTTCGTATTGCTATATACTGCGATACACGGTTACCTGTTTTGCAATACGTAA
ACTTAGGCAAATATGACAGTTTTTCCATGATTATTTTACGTTTCAATGCTTAAATGGT
CTTATCTGTTATCTCCTTGAAGGTTTAATAAAATAACAATAAACATAAATCATTATTGAA
AATTAATGAACAAAAGTAAAGCGCTTCTCAGTTACCTTAACCTAACTTATTTATGAATGG
GATTACTATCCAAGAATGTGAAATTACAAACACCTTGGGATAACACTGCAAAACGACTG
TTCATGGGACGGACATGAAAAAGGTGAGTCCCATGTTAAACTGTTGAGAAAGTTTCCTAT
ACTGTTTGTCCCGAAAAAGGCTAAAGACCATGTACTAATCAATTATTCTATCTATTTTCG
ATTACTGTTCTCATATTTGGGACAACGTGTGCAGATCGGTAGCATCCAAGCTCGTCTAAAT
CGGTTTGATAAACCTTGTCAAATAACATGTTGTCTCAACATCCAAGCTCACCTAAACCTT
GTCAATACCTGCATCTGAACAAATGTATATTTAAGACGATAGCATCCAAGCTCATCTTTA
AAATGAATATTTTCTCTTTTTCTACCAAAACATTATTTGGTTGACAGTTGTCCTCCCTAT
TATAGTAAAAAGAACTGGGTGGCAATTGTCCTAGGTGGCAATTGTCCGGATGGCAATTGT
CCGGGTGGCAATTGTCCGGGTGGCAGTTGTCCAGGTGGCTATTGTCCTGTTCCCATATTT
ACGTATCCCATTTTCTGCTCTGTAATTTTAAATAAACTCACCTGCCTAAGGTAAGACGAC
ATGTGTCACGTGAACATCGTTTGGGGGCAAGGGCGGAATCCCTTCGTTGAAAGTAAATGA
ATACTGTACATAGAGATGCGTATCTTGAACCTTTATTAGCTTTGATATTGTGCTTAATA
TTACATGAATGTATTTCAATATGTAATTATGTGTTCAAATGAATGGTTGACTTGAATGGT
TTTATTGCTTTATATGCTACATCAACATGTGTGTTTCTTTTCATTTGAG

DOMAIN 1B

CACCTGTGCATCATCGCCACGATGACGATCTTATTGTTGAAAAAATATAGATCATTGGA
CTCGTGAAGAGGAATACGAGCTAAGGATGGCTCTGGAGAGATTCCAGGCCGACACATCCG
TTGATGGGTACCAGGCTACAGTAGAGTACCATGGCCTTCCTGCTCGTTGTCCACGACCAG
ATGCAAAAGTCAGGTTTCGCCTGTTGTATGCATGGCATGGCATCCTTCCCTCACTGGCACC
GGCTGTTTCGTTACCCAGGTGGAAGATGCTCTTGTACGGCGTGGATCGCCTATCGGTGTTT
CTTATTGGGACTGGACAAAACCTATGACTCACCTTCCAGACTTGGCATCAAATGAGACGT
ACGTAGACCCGTATGGACATACACATCATAATCCATTCTTCAATGCAAATATATCTTTTG
AGGAGGGACACCATCACACGAGCAGGATGATAGATTGAACTGTTTGCCCCAGTCGCTT
TTGGGGAGCATTCCCATCTGTTTGATGGAATCCTGTACGCATTGAGCAGGAAGATTTCT
GCGACTTTGAGATTGAGTTTGTAGTTAGTCCATAATTCTATTGATGCGTGGATAGGCGGTT
CCGAAGATTACTCCATGGCCACCCTGCATTACACAGCCTTTGACCCCATTTTCTACCTTC
ATCATTCCAATGTCGATCGTCTATGGGCAATCTGGCAAGCTCTTCAAATCAGGAGACACA
AGCCATATCAAGCCCACTGTGCACAGTCTGTGGAACAGTTGCCAATGAAGCCATTTGCTT
TCCCATCACCTCTTAACAACAACGAGAAGACACATAGTCATTGAGTCCCGACTGACATTT
ATGACTACGAGGAAGTGCTGCACTACAGCTACGATGATCTAACGTTTGGTGGGATGAACC
TTGAAGAAATAGAAGAAGCTATACATCTCAGACAACAGCATGAACGAGTCTTCGCGGGAT
TTCTCCTTGCTGGAATAGGAACATCTGCACTTGTGACATTTTCATAAATAAACCGGGGA
ACCAACCACTCAAAGCTGGAGATATTGCCATTCTTGGTGGTGCCAAGGAAATGCCTTGGG
CGTTTGACCGCTTGTATAAGGTGCAAAATACTGACTCATTGAAGACACTTTCTCTCGATG
TCGATGGAGATTATGAAGTCACTTTTAAATTCATGATATGCACGGAAACGCTCTTGATA
CGGACCTGATTCCACACGCAGCAGTTGTTTCTGAGCCAGCTCACC

INTRON 1B/1C (SEQ ID NO:115)

GTAAGTAAATTTACAAAATTTGGTGTCTCTAACTATCCTAAGTATTCAATCGTTAGCGT
GTACCTATCTGCATAATGCAATACCCTGACTCCATATAAGTATAGTATATTTACTCTGGT
CGAAAACAAACAAATTGAAAACAAGAGTGGACGTGCTGTTATGATTTCTTTTTTCATTCTT
GGTTCGTTGTGTAATGCCACAGCCAGCAATTCAGATATATAGCGACGGTCTATGAATAC
TCCAGTCTGGACCAGACAATCGTGTGGAATGGTTTAGGCACATTATATCAAATTCATTGT

Fig. 4g

TGAAGATATGAGTTATGAGGTCACAATGTTGTCTTGTTACCCCGTGTCAGTAGTGACGTC
ATTTTCATGACTGAAATCTCTTCAACGCCGTTTAGCAATAATAGGCTCAGTAGTATTCAAC
CAATTACAATCAGTAGAAAATTCTCTATACTATTCTTATGTTGCATCCTGATATCCCTAT
GCAAAAATTAGTCATCTAATATAATCATTTCGATAAAATACTTTGGGCAAACAAATCAAT
GTAACATCTATTTTCTTTCAG

DOMAIN 1C

CTACCTTTGAGGATGAAAAGCACAGCTTACGAATCAGAAAAAATGTCGACAGCTTGACTC
CTGAAGAAACAAATGAACTGCGTAAAGCCCTGGAGCTTCTTGAAAATGATCATACTGCAG
GTGGATTCAATCAGCTTGGCGCCTTCCATGGAGAGCCTAAATGGTGCCCTAATCCTGAAG
CGGAGCACAAGGTTGCATGCTGTGTTTCATGGCATGGCTGTTTTCCCTCATTGGCACAGGC
TTCTTGCTCTCCAGGCGGAGAATGCTCTTAGAAAGCATGGGTACAGTGGTGCTCTACCAT
ACTGGGATTGGACTCGCCCCCTTTCCCAACTTCCTGATCTGGTTAGTCATGAGCAGTATA
CAGATCCTTCCGACCATCACGTGAAGCATAACCCGTGGTTCAATGGCCACATCGATACAG
TAAATCAGGATACCACCAGAAGCGTACGGGAGGATCTTATCAACAACCTGAATTTGGAC
ATTTACGGATATTGCTCAACAAGTCCTCTTAGCATTAGAACAAGATGACTTCTGTTCGT
TTGAAGTGCAGTATGAGATTTCCCATAAATTTATCCATGCACTTGTAGGAGGAACCGACG
CTTATGGCATGGCATCGCTGAGATATACAGCATACGATCCAATCTTTTTCTTGATCATT
CAAACACCGACAGGATCTGGGCTATTTGGCAATCCCTGCAAAAATACAGAGGCAAACCGT
ACAACACTGCCAACTGCGCCATAGAATCTATGAGAAGGCCCTGCAACCATTTGGACTAA
GCAGTGCCATTAACCTTGACAGAATCACCAGAGAGCATGCTATCCCGTTTGATGTCTTCA
ACTATAGAGATAACCTTCATTACGTATATGATACCCTGGAATTTAATGGTTTGTGCGATTT
CACAACCTTGATAGAGAGCTGGAAAAAATCAAGAGTCACGAAAGAGTATTTGCTGGATTCT
TGCTGTGCGGGATTAAAAAATCTGCTCTTGTGAAATTCGAAGTTTGTACTCCACCTGATA
ATTGTCATAAAGCAGGGGAGTTTTATCTACTCGGGGACGAAAACGAGATGGCTTGGGCCT
ATGACCGACTTTTCAAGTATGATATTACTCAGGTTCTGGAAGCAAACCATCTACACTTCT
ATGATCATCTCTTCATTTCGCTACGAAGTCTTTGATCTTAAAGGAGTGAGTTTGGGAACTG
ACCTGTTCCACACTGCAAATGTGGTACATGATTCCGGCACAG

INTRON 1C/1D (SEQ ID NO:116)

GTACGTGGATTTGATTACATAGCAATGCTATATGATTTAGTAATTACAACCTCAAGTCA
TGTAGCCGTTTTAGATTGCATTACATCAAACAGCATTGGATTAAATTGGGGGATTGTCCA
GGCCGCATTATGTTGCATTCCGAAAATAGTTTTGTGTCCAGTGTCCACGTTTAAATTTAA
CCATTTTAATCATATTAGGGATAATTTTAATAGATGTTATAGTGCTTTATTTTCATATTGT
TACAGTGGACAGTCACCAAGGACATATTTTACTCTATAGATACACAAACACCAATTAAAA
CCCTGCTTTGGAAAGTCTAACTTTTTCCCCACAG

DOMAIN 1D

GCACCCGTGATCGTGATAACTACGTTGAAGAAGTTACTGGGGCCAGTCATATCAGGAAGA
ATTTGAACGACCTCAATACCGGAGAAATGGAAAGCCTTAGAGCTGCTTTCCTGCATATTC
AGGACGACGGAACATATGAATCTATTGCCAGTACCATGGCAAACAGGCAAATGTCAAT
TGAATGATCATAATATTGCGTGTTGTGTCCATGGTATGCCTACCTTCCCCCAGTGGCACA
GACTGTATGTGGTTTCAAGTGGAGAATGCTCTCCTAAACAGGGGATCTGGTGTGGCTGTTC
CTTACTGGGAGTGGACTGCTCCCATAGACCATCTACCTCATTTCATTGATGATGCAACAT
ACTTCAATTCCCGACAACAGCGGTACGACCCTAACCCTTTCTTCAGGGGAAAGGTTACTT
TTGAAAACGCAGTCACAACAAGGGACCCACAAGCCGGGCTCTTCAACTCAGATTATATGT
ATGAGAATGTTTTACTTGCACTGGAGCAGGAAAATTATTGTGACTTTGAAATTCAGTTTG
AGCTTGTTTCATAACGCACTTCATTCCATGCTGGGAGGTAAAGGGCAGTACTCCATGTCTCT

Fig. 4h

CCCTGGACTATTCTGCGTTTGATCCCGTCTTCTTCCTACATCATGCCAACACGGACAGAC
TGTGGGCAATCTGGCAGGAACTACAAAGATTCCGAGAACTGCCTTATGAAGAAGCGAACT
GTGCAATCAACCTCATGCATCAACCACTGAAGCCGTTTCAGTGATCCACATGAGAATCACG
ACAATGTCACTTTGAAATACTCAAACCACAGGACGGATTCGACTACCAGAACCACTTCG
GATACAAGTATGACAACCTTGAGTTCCATCACTTATCTATCCCAAGTCTTGATGCTACCC
TGAAGCAAAGGAGAAATCACGACAGAGTGTTTGCGGGCTTCCTTCTTCATAACATAGGAA
CTTCTGCTGACATAACTATCTACATATGTCTGCCTGACGGACGGCGTGGCAATGACTGCA
GTCATGAGGCGGGAACATTCTATATCCTCGGAGGCGAAACAGAGATGCCTTTTATCTTTG
ACCGTTTGTATAAATTTGAAATCACCAAACCACTGCAACAGTTAGGAGTCAAGCTGCATG
GTGGAGTTTTTCGAACTGGAGCTTGAGATCAAGGCATACAACGGTTCCTATCTGGATCCCC
ATACCTTTGATCCAACTATCATCTTTGAACCTGGAACAG

INTRON 1D/1E (SEQ ID NO:117)

GTAATGCCATCTTAATACAGTTCGTTTCGTTAAATTATATATGTTCGTTTACAACACCATA
CCTTGAATTGAGGTAATACATCACTTGATATTGATAATGTAATGGTAATTGTTCTTGTTT
GTAAAACCGTTTCTGGGGTGTTTATTCATCTACCTGGTGGATAGTGAGTAAACACAT
TCGGTTTAATATGGGTATCTAATGGACAGTGAAGTGTGCTGGCTAGGCAGATACCTTGGT
TTCTGTGAATGGAGGTAGTAGAAAGGGGTTTTGATGATTGCAG

DOMAIN 1E

ATACCCATATCTTGGACCACGACCATGAGGAAGAGATACTTGTGAGGAAGAATATAATTG
ATTTGAGCCCAAGGGAGAGGGTTTCTCTAGTCAAAGCTTTGCAAAGAATGAAGAATGATC
GCTCCGCTGATGGGTACCAAGCCATTGCCTCTTTCCATGCCCTGCCACCACTCTGTCCCA
ATCCATCTGCAGCTCACCGTTATGCTTGCTGTGTCCATGGCATGGCTACATTTCCCCAGT
GGCACAGACTGTACACTGTTTCAGGTTTCAGGATGCCCTGAGGAGACATGGTTCACTTGTTG
GTATTCCTTACTGGGACTGGACAAAACCAGTCAACGAGTTACCCGAGCTTCTTTCTTCAG
CAACATTTTATCATCCAATCCGGAATATTAATATTTCAAATCCATTCCCTCGGGGCTGACA
TAGAATTTGAAGGACCGGGCGTTTCATACAGAGAGGCACATAAATACTGAGCGCCTGTTTC
ACAGTGGGGATCATGACGGATACCACAACCTGGTTCTTCGAACTGTTCTCTTTGCTTTGG
AACAGGAAGATTACTGCGATTTTGAATAACAATTTGAGATAGCCCATAATGGCATCCACA
CATGGATTGGTGGAAGCGCAGTATATGGCATGGGACACCTTCACTATGCATCATATGATC
CAATTTTCTACATCCACCATTACAGACGGACAGAATATGGGCTATTTGGCAAGAGCTGC
AGAAGTACAGGGTCTATCTGGTTTCGGAAGCAAACCTGTGCCATTGAACATATGAGAACAC
CCTTGAAGCCTTTCAGCTTTGGGGCCACCCTACAATTTGAATAGTCATACGCAAGAATATT
CAAAGCCTGAGGACACGTTTGAATAAAGAAGTTTGGATACAGATATGATAGTCTGGAAT
TGGAGGGGCGATCAATTTCTCGCATTGATGAACTTATCCAGCAGAGACAGGAGAAAGACA
GAACTTTTGCAGGGTTCCTCCTTAAAGGTTTTTGGTACATCCGCATCTGTGTCAATTGCAAG
TTTGCAGAGTTGATCACACCTGTAAAGATGCGGGCTATTTCACTATTCTGGGAGGATCAG
CCGAAATGCCATGGGCATTTCGACAGGCTTTATAAGTATGACATTACTAAACTCTTCACG
ACATGAACCTGAGGCACGAGGACACTTTCTCTATAGACGTAACATACAGTCTTACAATG
GAACAGTACTCTCGGGAGACCTCATTACAGACGCCCTCCATTATATTTGTACCTGGACGCC

INTRON 1E/1F-1 (SEQ ID NO:118)

GTGAGTACCTGTTTGCCTAAGACTTCTGTAGGCTAAAAGTGTAAGAAATATCAATTAAT
TTCAATTCACCCAAACTTGAAAACGGTACCTATATAGGTAACTTTTTGTCTACAGTAA
CTGAACATACCTACACATTTTCATGAAATGATCTCTCAATATTTTCCACCAACAG

Fig. 4i

DOMAIN 1F-1 (1st part of domain f)

ATAAACTCAACTCACGGAAACATACACCTAACAGAGTCCGCCATGAGCTAAGTAGCCTTA
GTTCCCGTGACATAGCAAGCTTGAAGGCAGCTTTGACAAGCCTTCAACATGATAATGGGA
CTGATGGTTATCAAGCTATTGCTGCCTTCCATGGCGTTCCTGCGCAGTGCCACGAGCCAT
CTGGACGTGAG

INTRON 1F-1/1F-2 (SEQ ID NO:119)

GTAAATTTACAGAGCTTTATGAAGTGTGTTTCAGAGTGAAGAGACCAAGATATACTTATAC
CCAAAACCTAGCTAGCAACAGACGATTTCACTTGTTCGGACACTTTGTATTATACGTTGG
ATCCCAAGGTAAACGGAAACGTAACCGAGAATCAGTCCGTAAAGTGAGTGAGTGAGTTTG
GGGCTTAACGTCGCACTCAGCAATACCCAGCTATGTGGCGACTCTCAGATTTACTGCTG
GAGGAGAACCTACATAGCCCGGTTTAACCCGTGTGGTATGTAGTAAGACCAGCGCGGCAT
GGCTGGTATCTGACGGACGAAGGGTGGCGCTGCACGTATTCCAGTGGTACAACACTGCAC
CCCAATTTACCCGACCGGAGAAGTATCTCCCTTCGGAGATATCGCTGCCTTCCACGG
GATTCGAACTCGGTGACCTTCAAGCCAGCGCGCTTCTAGCGGGGGCGATTAGAGGTTNAA
GGCCGACGGCTCTACCACCTTAACCTATCCCCCGGCCCACTCCTGACGGAAATGTTTATA
ATTGAGCCTTTGTTTTCTTATTAACACTCTTGGCAGATTTTCTATAGATAATGGATTCA
CATGTAGACAGTCTCCCATTTGTTGTAAGTGGTAGTCAAGAGTTAGAATCTGAATACATTC
TCCAAGATGGATCAAGGAAAACAATAATTACTTGATGTTGCAG

DOMAIN 1F-2 (2nd part of domain f)

ATCGCCTGTTGCATCCACGGCATGGCGACGTTTCCTCACTGGCACCGGTTGTACACTCTG
CAGTTGGAGCAAGCGCTGCGCAGACACGGGTCCAGTGTGCTGTTCCATACTGGGACTGG
ACCAAGCCAATCACCGAACTGCCACACATTCTGACAGACGGAGAATATTATGACGTTTGG
CAAAATGCCGTCTTGGCCAATCCGTTTGCAAGAGGTTATGTGAAAATTAAAGATGCATTT
ACGGTGAGAAATGTCCAGGAAAGTCTGTTCAAAATGTCAAGTTTTGGAAAGCACTCGCTT
CTGTTTGACCAGGCTTTGTTGGCTCTTGAACAACTGACTACTGTGACTTCGAAGTTCAG
TTTGAAGTGATGCATAACACGATCCATTATCTCGTAGGAGGGCGTCAAACGTACGCCTTC
TCCTCTCTCGAGTATTCCTCATAACGATCCAATCTTCTTTATTACCACTCGTTTGTGAC
AAAATATGGGCTGTATGGCAAGAACTGCAAAGCAGGAGACATCTACAGTTTAGAACAGCT
GATTGTGCTGTGGGCCTCATGGGTGAGGCAATGAGGCCTTTCAACAAGGATTTCAACCAC
AACTCGTTACACCAAGAAGCACGCAGTCCCTAATACAGTATTTGATTATGAAGATCTTGGC
TATAACTATGACAACCTTGAAATCAGTGGTTTAACTTAAATGAGATCGAGGCGTTAATA
GCAAAACGCAAGTCACATGCTAGAGTCTTTGCTGGGTTCTGTTGTTTGGATTAGGAACT
TCGGCTGATATACATCTGGAAATTTGCAAGACATCGGAAAACCTGCCATGATGCTGGTGTG
ATTTTCATCCTTGGAGGTTCTGCAGAGATGCATTGGGCATACAACCGCCTCTACAAGTAT
GACATTACAGAAGCATTGCAGGAATTTGACATCAACCCTGAAGATGTTTTCCATGCTGAT
GAACCATTTTTCTGAGGCTGTGCGTTGTTGCTGTGAATGGAAGTGCATTCCATCGTCT
CATCTTCACCAGCCAACGATAATCTATGAACCAGGCGAAG

INTRON 1F-2/1G-1 (SEQ ID NO:120)

GTGAGATATATGCAAATTGAATGTTGTCCAGATGCGTTGTTTACATTTATATGCTTGGAA
TTGTCCTGAACGAATACAGTGGAAATAACCAAAAGCTGAAAAATAAAAAGATATATACTTC
ATTCTGAATTTGTGAGTATTGCTGACCCAAAACACGTTATCCATGTGACACTATATTT
GCCTTTCTGAATCTGAGACTGCGTTATGTTTCTAATAATCACGAAATATGGTATACAGGT
TGTGTATCTGTAGAATACCCAAGGCAGAAATTTAAAGGGTCACACCCTGTTTAATACAG

Fig. 4j

DOMAIN 1G-1 (1st part of domain g)

ATCACCATGACGACCATCAGTCGGGAAGCATAGCAGGATCCGGGGTCCGCAAGGACGTGA
ACACCTTGACTAAGGCTGAGACCGACAACCTGAGGGAGGCGCTGTGGGGTGTTCATGGCAG
ACCACGGTCCCAATGGCTTTCAAGCTATTGCTGCTTTCCATGGAAAACCAGCTTTGTGTG
CCATGCCTGATGGCCACAACACTCATGTTGTACTCACG

INTRON 1G-1/1G-2 (SEQ ID NO:121)

GTAAGTTTGTGTTGGTTAGTGTGTTGGTTGCATGTTTTGCCATATCGATAGTATCAGTGTGG
TAACATCTGGTTTCTAGTTCATTTCAGTTCACCTTATCAGAAGCTGTTTGCTCTCGTCTAC
AATAGTGACGTCTTTTCAGTTTTAGAACCGTGTACATCCGGGTATATTGGTCTCCAGCAA
CCCGTGCTTGTCGTGGGAGGCCACTGATGGGAACGGGTGGTCAGACTCGCTCACTTAGTT
GACACATGTCAATTGCGAAGATCGATGCTGAGGTGTAAACATTGGATTGTCTGGTCCA
GACTCGATTATTTACAGACAGCCGCCATGTACCTGGAATATTGCTGAGTGCGGCGTTAA
CAACAACTAGTCAGACTAATCTTTCACTGTTTTATAATGATGGCTCGAACCTAGCACTCA
TGTCCCAAGTTGGCGAACATCTGGAAGGGAATTTCAAATGAAAAGAACAATCTTTCACGT
CTATTGGTATCACGCTCCTGGAGAAGAACATGATGTTACGGCGTTACTTCCTCTTACCT
GTTTTACTTGTTCCACGTTTCTTCATATTTAAAGAGTATTTGGGTATTAGAGCTTTGGT
GCTGTTACAATGCTACTCAACTGTTTCAGTGCGGGCGACCGCGCTTGTTTACACATTAAGT
TTTGTTTGTGTTGGTTGGTTTGTGTGTGTGTGTGTATGTGTGTGTGTGTGTGTGTGTGTA
TGTGTGTGTGTGTGTATCTATGTCTATGTGTCTGTGTCTGTGTGTCTGTCTATGTGTGTG
TGTGTCTGTGTCTATGTGTGTGTCTGCGTGTGTGTCTGTGTCCGTATGTGGCTGTGTCTA
TGTGTGTGTGTGTCTGTGTTTATGTGTGTATATGCGTGTGTGTCTGTGTCCGTATGTGGC
TGTGTCTATGTGTGTGACATGCAATACATGCTGTGATACTCACTAGCTGCGTCTATCGAC
CAG

DOMAIN 1G-2 (2nd part of domain g)

GCATGGCTACCTTCCCACACTGGCATCGCCTCTACACCAAGCAGATGGAGGATGCAATGA
GGGCGCATGGGTCTCATGTCGGCCTGCCCTACTGGGACTGGACTGCTGCCTTCACCCACC
TGCCAACACTGGTCACCGACACGGACAACAACCCCTTCCAACAT

INTRON 1G-2/1G-3 (SEQ ID NO:122)

GTAAGAGCGGGGTAGGGATGGGGTGGTAGGGGGTGGGTGTTCTATTACTTCCCGCTTCA
CTTGATGAAATGGATAACCTTGGCTGCATCCCAATTGCGTGATCGATTCTCTTTCGATT
CACTCGTGCGATTAGACTGCCTTATTTACTATAGTAGTTAGAATGTTGCTCAGTGCGCCG
TTAAACAATAACACAAAACCGCATTTGTTTTATATGGTCACTCTACTGTTTATCACG
TATATGTATGTTCCGACTCACTGGTTGGTGCGTACCATTCTACTGTCACTGAGAGCCA
ATGTTCTCAGATGTGTGAAATGTTTGAAAGCCGTTTCTACATAATATTGCAGGAATACCA
TTGTAGAATGTAGTCAAACAGGTAACAATCTGTTAGTGAGCCCAGTTCGAGGTTGCGTTG
TAGGGTGTAGTCCAACAGGTAGGCAGTCCATAAGCATAGTTTTTAAGCATTTTAGATCAT
CTATAATTAACCACATGGTTAGCCGCTATGTTTAGTTTAATCCAGTATAAGTTAGAAGTG
TTATATTTTCGAAGGGAAGTGAGTAAATCCTTATTCCTTGACTACCATTTAATAGATTTC
CAATGACTCCATTCAACTCCTAACTTTACATCACTGCTCTCTTCAACAG

DOMAIN 1G-3 (3rd part of domain g)

GGACACATTGATTATCTCAATGTCAGCACAACTCGATCTCCCCGAGACATGCTGTTCAAC
GACCCCGAGCATGGATCAGAGTCGTTCTTCTACAGACAAGTCCTCTTAGCTCTGGAACAA

Fig. 4k

ACTGATTTCTGCAAATTCGAAGTTCAGTTTGAGATAACCCACAATGCCATCCATTCTGG
ACAGGTGGCCACAGCCCCTACGGAATGTCCACTCTCGACTTCACTGCCTACGATCCTCTC
TTCTGGCTTCACCACTCCAACACCGACAGAATCTGGGCTGTCTGGCAAGCTTTGCAAGAA
TACAGAGGACTTCCATACAACCATGCCAATTGTGAGATCCAGGCAATGAAAACGCCCCTG
AGGCCTTTTCAGTGACGATATCAACCACAACCCAGTCACAAAGGCTAACGCGAAGCCATTA
GATGTGTTTCGAGTATAATCGGTTGAGCTTCCAGTACGACAACCTCATCTTCCATGGATAC
AGTATTCCGGAACCTTGATCGCGTGCTTGAAGAAAGAAAGGAGGAGGACAGAATATTTGCT
GCCTTCCTTCTCAGTGGAATCAAGCGTAGTGCTGATGTAGTGTTTCGACATATGCCAGCCA
GAACACGAATGTGTGTTTCGAGGGACTTTTTCGATTTTGGGAGGGGAGCTAGAAATGCCC
TGGTCCTTCGACAGACTGTTCCGCTATGATATCACCAAGGTGATGAAGCAGCTACACCTG
AGGCATGACTCTGACTTTACCTTCAGGGTGAAGATTGTCGGCACCGACGACCACGAGCTT
CCTTCAGACAGTGTCAAAGCACCAACTATTGAATTTGAACCGGGCG

INTRON 1G-3/1H (SEQ ID NO:123)

GTGAGTACGACAGGCATTTCTAGTAAAAACCTACTTTTGGTAAAAGGTTTCGAGAAATCAC
TTGAAGCAACAACATGATTTTGTAAACGCCTATTACACGTGAACATGTCACACCCGGTGAT
GCCGTTTAATGGACATGCCTCTGTTAATGAAAGGGGTAAAGTACATGTGTATGGGGATGGG
ATGGGAGCCACCTGTCCCAATTTTCATAGGTCCCTAGGATCCCAGTTGCGTAGGAATCCCC
TGATTAATGCCTTGTGAATTCCTCCTGGAATTGTCTTGGCCCAAATTTTACAAACCCGC
CCCGATATACCTTGGAAATAATTGGGCCTAAGGGTGGGGCTTTTAAGGACCAAGAACCCA
ACCTAAACCCCAACCCATTTTTTCCACCCATTCCAGGTTTTGTTTTACCAAATAAAAAG
GTTTCCACTTTGAGGAAACCCTTTAAGGGTCTTTTCAGGGCTTTTTTTCTTTTCTGGGA
ATTCCAATTCCGGGGGAACAAAATACATATATTTACAGACCTTTGGTCAAATTTATATA
ATTTCCGACTTCATGTCATAGGTTTGTCTTTCTTCTTACACAG

DOMAIN 1H

TGCACAGAGGCGGAAACCACGAAGATGAACACCATGATGACAGACTCGCAGATGTCCTGA
TCAGGAAAGAAGTTGACTTCCTCTCCCTGCAAGAGGCCAACGCAATTAAGGATGCACTGT
ACAAGCTCCAGAATGACGACAGTAAAGGGGGCTTTGAGGCCATAGCTGGCTATCACGGGT
ATCCTAATATGTGTCCAGAAAGAGGTACCGACAAGTATCCCTGCTGTGTCCACGGAATGC
CCGTGTTCCCCCACTGGCACC GCCTGCATACCATTGAGATGGAGAGAGCTCTGAAAAACC
ATGGCTCTCCAATGGGCATTCTTACTGGGATTGGACAAAGAAGATGTGAGTCTTCCAT
CTTTCTTTGGAGATTCCAGCAACAACAACCTTTCTACAAATATTACATCCGGGGCGTGC
AGCACGAAACAACCAGGGACATTAATCAGAGACTCTTTAATCAAACCAAGTTTGGTGAAT
TTGATTACCTATATTACCTAACTCTGCAAGTCTTGAGGAAAACTCGTACTGTGACTTTG
AAGTTCAGTATGAGATCCTCCATAACGCCGTCCACTCCTGGCTTGGAGGAACTGGAAAGT
ATTCCATGTCTACCCTGGAGCATTGCGCCTTTGACCCTGTCTTCATGATTACCACTCGA
GTTTGGATAGAATCTGGATCCTTTGGCAGAAGTTGCAAAAGATAAGAATGAAGCCTTACT
ACGCATTGGATTGTGCTGGCGACAGACTTATGAAAGACCCCTGCATCCCTTCAACTACG
AAACCGTTAATGAAGATGAATTCACCCGCATCAACTCTTTCCCAAGCATACTGTTTGACC
ACTACAGGTTCAACTATGAATACGATAACATGAGAATCAGGGGTCAGGACATACATGAAC
TTGAAGAGGTAATTCAGGAATTAAGAAACAAAGATCGCATATTTGCTGGTTTTGTTTTGT
CGGGCTTACGGATATCAGCTACAGTGAAAGTATTCATTTCATTTCGAAAACGATACAAGTC
ACGAAGAATATGCAGGAGAATTTGCAGTTTTGGGAGGTGAGAAGGAGATGCCGTGGGCAT
ATGAAAGAATGCTGAAATTTGGACATCTCCGATGCTGTACACAAGCTTCACGTGAAAGATG
AAGACATCCGTTTTAGAGTGGTTGTTACTGCCTACAACGGTGACGTTGTTACCACCAGGC
TGTCTCAGCCATTCATCGTCCACCGTCCAGCCCATGTGGCTCACGACATCTTGGTAATCC
CAGTAGGTGCGGGCCATGACCTTCCGCCTAAAGTCGTAGTAAAGAGCGGCACCAAAGTCG

Fig. 4I

AGTTTACACCAATAGATTTCGTCGGTGAACAAAGCAATGGTGGAGCTGGGCAGCTATACTG
CTATGGCTAAATGCATCGTTCCCCCTTTCTCTTACCACGGCTTTGAACTGGACAAAGTCT
ACAGCGTCGATCACGGAGACTACTACATTGCTGCAGGTACCCACGCGTTGTGTGAGCAGA
ACCTCAGGCTCCACATCCACGTGGAACACGAGTAG

3'UTR

TTCACAG

INTRON 3'UTR (SEQ ID NO:124)

GTGAGGAGAAGGCCCCAGGCTAGCAGGGCAATGGATGAAGGAAATAGGGGCAAAGGGAAT
AGCAGTTACACCATCGACATTTCCAACCTCCTCAGAACTAATATATAGCCTTAATACAA
CCAGCCAAGACTCAACGGGCAGCCGGGGTGGGGGGATTGGTGGTTCGCTGTTTCAGACCA
GGGTGCAAATATCAGTGCGCAAATCAACATGTTGCGTGTGACACACTGACACAGCAGTC
ATTGAACCTGCAGACCCATAACAGGAAAATGGGGCAGATACGATCAAAGACAGTGTA
TAGGGATAAGTAGGCATATGCAACCACCTGATGGAAATGAAAAGGGGTAAGTTTAAACCC
CGGCTACCAAAGGTCCAATGGTTCCTTAACCCAGCTTACGCTATCCCTCTAATTTTCAGTA
TTGAGCTGATTTCTGTGAGTTCATGTAACTGTATACTTTCTGTATTATTACAG

3'UTR

GTTGCTATGCCGACTGCGCTATATTGGTGAACGAGACGATGAGGACATCTCTGAAAGAGT
TCGCCAAGTGATGTGTAGGTCACGGAAGTATTGTTGAGCTAACAATATGATGATTTCAA
ATGACTTGGCGCTCTAGGACAAAGACATAATTCATCAGCACCCCTGTGCACCAACTCTTTG
TTTGCTGCAAACGTCTGACAAGCGACACGTCAATCAACAAGCTGTTCAAACCTCAAGTGGA
TGTAAGTAGAATCGTTGGGCCATCGTTCACAAAGTATTGACAGATGTCACACATGATGGC
GAGAAACACTTTTAGAACTTTTAATGACCTAGAGTGACTTGTAATATGTAAATATATTCT
TCAAAGACTCAGCTGAACTATTGTTGGATAACACATCAATTCCTCAACAAAATGCTTTA
TCTTCACATGGATGTATGTAATGTGGCCGGCAATAAAGTATATATATGTAT

Fig. 5aPrimary structure of the HtH1 proteinSIGNAL PEPTIDE

LVQFLLVALVVGAGA

DOMAIN A

DNVVRKDVSHLTVDEVQALHGALHDVTASTGPLSFEDITSYHAAPASCDYKGRKIACCVHGMPSFP
FWHRAVYVQAERALLSKRKTVMGPYWDWTQTLTHLPSLVTEPIYIDSKGGKAQTNYWYRGEIAFIN
KKTARAVDDRLFEEKVEPGHYTHLMETVLDALQDEFCKFEIQFELAHNAIHVLVGGKFEYSMSNLE
YTSYDPIFFLHHSNVDRLFIAIWQRLQELRGKNPNAMDCAHELAHQQLQPFNRDSNPVQLTKDHSTP
ADLFDYKQLGYSYDSLNLNGMTPEQLKTELDERHSSKERAFAFASFRLSGFGGSANVVVYACVPDDDP
SDDYCEKAGDFFILGGQSEMPWRFYRPFYDVTEAVHHLGVPLSGHYVVKTELFVNGTALSPDLL
PQPTVAYRPGK

DOMAIN B

GHLDPVHHRHDDDLIVRKNIDHLTREEEYELRMALERFQADTSVDGYQATVEYHGLPARCPRPDA
KVRFACCMHGMAFPHWHRLFVTQVEDALVRRGSPIGVPYWDWTKPMTHLPDLASNETYVDPYGH
HHNPPFNANISFEEGHHTSRMIDSKLFAPVAFGEHSHLFDGILYAFEQEDFCDFEIQFELVHNSI
HAWIGGSEYDYSMATLHYTAFDPIFYLHHSNVDRLWAIWQALQIRRHKPYQAHCAQSVEQLPMKPFA
FPSPLNNEKTHSHSVPTDIYDYEEVLHYSYDDLTFGGMNLEIEEAIHLRQQHERVFAGFLLAGI
GTSALVDIFINKPGNQPLKAGDIAILGGAKEMPWAFDRLYKVEITDSLKTLSLDVDGDYEVTFKIH
DMHGNAALDLDLIPHAADVSEPAH

DOMAIN C

PTFEDEKHSLRIRKNVDSLTPREETNELRKALELLENDHTAGGFNQLGAFHGEPKWCPNPEAEHKVA
CCVHGMVFPWHRLALQAENALRKHGYSALPYWDWTRPLSQLPDLVSHEQYTDPSDHVVKHNP
WFNGHIDTVNQDTTRSVREDLYQQPEFGHFTDIAQQVLLALEQDDFCSEVQYEISHNFIHALVGG
TDAYGMASLRYTAYDPIFFLHHSNTDRIWAIWQSLQKYRGKPYNTANCAIESMRPLQPFGLSSAI
NPDRITREHAIPFDVFNRYRDNLHYVYDTLEFNGLSISQLDRELEKIKSHERVFAGFLLSGIKKSAL
VKFEVCTPPDNCHKAGEFYLLGDENEMAWAYDRLFKYDITQVLEANHLHFYDHLFIRYEVFDLKG
SLGTDLFHTANVVHDSGT

DOMAIN D

GTRDRDNYVEEVTGASHIRKNLNDLNTGEMESLRAAFLHIQDDGTYESIAQYHGKPGKCQLNDHNI
ACCVHGMPTFPQWHRLYVVQVENALLNRGSGVAVPYWEWTAPIDHLPHFIDDATYFNSRQQRYDPN
PFRGKVTFFENAVTTRDPQAGLFNSDYMYENVLLALEQENYCDFEIQFELVHNALHSMGLGKGQYS
MSSLDYSAFDPVFFLHHANTDRLWAIWQELQRFRELPYEEANCAINLMHQPLKPFSDPHENHDNVT
LKYSKPQDGFYQNHFGYKYDNLEFHLSIPSLDATLKQRRNHDRVFAGFLLHNIGTSADITIYIC
LPDGRGNDCSHEAGTFYILGGETEMPFIFDRLYKFEITKPLQQLGVKLHGGVFEELELEIKAYNGS
YLDPHTFDPTIIFEPGT

DOMAIN E

DTHILDHDHEEEILVRKNIIDLSRPRVSLVKALQRMKNDRSADGYQAIASFHALPPLCPNPSAAH
RYACCVHGMATFPQWHRLYTVQVQDALRRHGSVLGIPYWDWTKPVNELPELLSSATFYHPIRNINI
SNPFLGADIEFEGPGVHTERHINTERLFHSGDHDGYHNWFFETVLFALQEDYCDFEIQFEIAHNG

Fig. 5b

IHTWIGGS AVYGMGHLHYASYDPIFYIHHSQTDRIWAIWQELQKYRGLSGSEANCAIEHMRTPLKP
FSFGPPYNLNSHTQEYSKPEDTFDYKKFGYRYDSLELEGRSISRIDELIQQRQEKDRTFAGFLLKG
FGTSASVSLQVCRVDHTCKDAGYFTILGSSAEMPWAFDRLYKYDITKTLHDMNLRHEDTFSIDVTI
TSYNGTVLSGDLIQTPSII FVPGR

DOMAIN F

HKLNSRKHTPNRVRHELSSLSSRDIA SLKAALTSLOHDNGTDGYQAIAAFHGVPAQCHEPSGREIA
CCIHGMATFPHWHRLYTLQLEQALRRHGSSVAVPYWDWTKPITELPHILTDGEYYDVWQNAVLANP
FARGYVKIKDAFTVRNVQESL FKMSSFGKHSLLFDQALLALEQTDYCDFEVQFEVMHNTIHYLVGG
RQTYAFSSLEYSSYDPIFFIHHSFVDKIWAVWQELQSRRLQFRTADCAVGLMGQAMRPFNKDFNH
NSFTKKHAVPNTVFDYEDLGYNYNLEISGLNLNEIEALIAKRKSHARVFAGFLLFGLGTSADIHL
EICKTSENCHDAGVIFILGSSAEMHWAYNRLYKYDITEALQEFDINPEDVFHADEPFFLRLSVVAV
NGTVIPSSHLHQPTIIYEPGE

DOMAIN G

DHHDDHQSGS IAGSGVRKDVNTLTKAETDNLREALWGMADHGPNGFQAIAAFHGKPALCPMPDGH
NYSCTHGMATFPHWHRLYTKQMEDAMRAHGSHVGLPYWDWTAFTHLPTLVTDTDNNPFQHGHI
YLVNSTTRSPRDMLENDPEHGSESFYRQVLLALEQTD FCKFEVQFEITHNAIHSWTGGHSPYGMS
TLDFTAYDPLFWLHHSNTDRIWAVWQALQEYRGLPYNHANCEIQAMKTPLRPFSD DINHPVTKAN
AKPLDVFEYNRLSFQYDNLI FHGYSI PELDRVLEERKEEDRIFAAFLLSGIKRSADVVDICQPEH
ECVFAGTFAILGGELEMPWSFDRLFRYDITKVMKQLHLRHDSDFTRVKIVGTDDHELPSDSVKAP
TIEFEPG

DOMAIN H

VHRGGNHEDEHDDRLADVLRKEVD FLSLQEANAIKDALYKLQND DSKGGFEAIAGYHGYPNMCP
ERGTDKYPCCVHGMPVFPHWHRLHTIQMERALKNHGSPMGI PYWDWTKKMSSLPSFFGDSSNNNPF
YKYYIRGVQHETTRDINQRLFNQTKFGEFDYLYLT LQVLEENSYCDFEVQYEILHNAVH SWLGGT
GKYSMSTLEHSADFDPVFMIIHSSLDRIWILWQKLQKIRMKPYALDCAGDRLMKDPLHPFNYETVN
EDEFTRINSFPSILFDHYRFNYEYDNMRIRGQDIHELEEV IQELRNKDRI FAGFVLSGLRISATVK
VFIHSKNDTSHEEYAGEFAVLGGEKEMPWAYERMLKLDISDAVHKLHVKDEDIRFRVVVTAYNGDV
VTTRLSQPFIVHRPAHVAHDILVIPVGAGHDLPPKV VVKS GTKVEFTPIDSSVNKAMVELGSYTAM
AKCIVPPFSYHGFELDKVYSVDHGDYIIAAGTHALCEQNLRLHIHVEHE

DOMAIN 2A-1 (1st part of domain a)
[domain a, parts 1-4: SEQ ID NO:156]

INTRON 2A-1/2A-2 (SEQ ID NO:125)

DOMAIN 2A-2 (2nd part of domain a)

INTRON 2A-2/2A-3 (SEQ ID NO:126)

GTGAGTCACGTTCTCTGATGGTCACGAGTCACGTTCTCTGATGGTCACGAGTCACGTTCTCTGATG
GTCACGAGTCACGTTCTCTGATGGTCACGAGTCACATTCTCTGATGGTCACGAGTCACATTCTCTG
TTGAGTGAAGTCTCAGTACCATTATTTCTCTTACCTTCTTCTAACCAGGGGTTTCAGCGTGGATC
GTCTGAGAAGTTAGCGCAAATCTATATTGAAGTCATTTTTCTATCATATAACCATCGTTATATCCA
CGTGCGAAAGTGTTCAATTAATTATTTTTATTTTCATTTATGAAGGTCTAAAAGAAAATATGTATTG
TTGGAAACTATATTCGAAGGTGAAGGCAACACGAGTGTATTAATATTCTCAATATCAATGTACGCT
CTGTCTAGCACCTGTTTCACCAGGAACACACCTTTAGCGTACCAAAATATCAGCTGATGATTTCTGA
AGCGGACTATACCCTCACCACCTTGTTTTGTGTGTGTATTTATGTGTGCATGTGTGTGCGTGCGTGC
GTGTGTGTGTGTGTCCTACGTATGTTGATATTTTGTCTGACTGTATATGTTTCGTGCTTACCATTG
AAG

Fig. 6b

DOMAIN 2A-3 (3rd part of domain a)

GTACTCCATGTCTCATCTCGAGTACACCTCCTACGACCCCCCTCTTCTTCCATCACTCCAACAC
CGACCGCATCTTCGCCATCTGGCAACGTCTTCAGGTACTCAGAGGAAAGGACCCCAACACCGCCGA
CTGCGCACACAACCTCATCCATGAGCCCATGGAACCGTTCCGTCGGGACTCGAACCCCTCTTGACCT
CACCAGGGAAAACCTCAAACCAATTGACAGCTTTGATTATGCCACCTTGGCTACCA

INTRON 2A-3/2A-4 (SEQ ID NO:127)

GTATGTATGATTCTAATAATGAATGTTTTTACCTCCGGTTTAAACAATATTTTAGTATTACGAAAG
GAGAAGTACCTCGAGAGGTCTAGGTCTCAGATGTTTAGAAACCCATGAAGACAGGTATGCTTCTGA
AAAACAAAGTAACATCATGAGGCTAAAGTTCAGATTCAAACCATCGTAGTTCGAATCCAGCATGCA
AAGGGCCCTAACCCTGTAGATGGCGCTGCTTGAAACAGAGTAGTCTGTTTCAGGGTCAGTACTGTCC
CCACAAACATCATAGTCAGGGTCAGTACTGTCCCCACAAACATCATAGTCAGGGTCAGTACTGTCC
CCACAAACATCACAGTCAGGGTTAATTTTGGATTTCGGTTTTCGAATGCGAAGAAGACAGTCACGCCC
TGACACTGGACCGAGGTTGCCGAGAAAGCTCGTGATATTGCTGGAATACTGCCCAGTAAAACCATC
ATTTATTTTtaggctATTTATTACGAAAAATAATAATATGTATAGAAATGCATATGATCGCTGTTTG
AATGTAAATTTAGAATGGGTTTGGGAGTGTTCACTATTTTTTTCATCAAAATTTTCATGTATTTTAA
CCGATCGACGCTGAAGACAAACTACCGTTAATCAGGCAGTTCATTTCATATCTGATAGGGAATATTG
GTTGTTAACCAACGCTACATTGTGTCCAG

DOMAIN 2A-4 (4th part of domain a)

GTATGATGACTTGACCCTGAACGGTATGACCCCAGAGGAATTGAACTCATATCTGCATGAACGGTC
AGGCAAGGAGGGGGTGTTCGCAAGCTTCCGACTCTCAGGTTTTGGCGGCTCTGCTAACGTTGTTGT
CTACGCATGCCGTCTGCCCACGATGAAATGGCTGTGATCAGTGCGACAAAGCCGGCGACTTCTT
TGTGTTGGGCGGACCCACCGAGATGCCCTGGAGGTTTTACAGAGCATTCCACTTCGACGTCACCGA
CAGCATCGACAACATCGACAAGGACCGCCACGGCCACTATTATGTAAAGGCGGAATTATTCAGTGT
AAATGGAAGTGCCTACCGAATGATCTCCTGCCTCAACCCACCATCTCACACAGGCCAGCCCGCGG
ACACGTTGATG

INTRON 2A-4/2B (SEQ ID NO:128)

GTAAATGGCCATTGTATACATGCATTCATTTGGACTTTGAGTGAGTGAGTGAGTGATGCGTATTCAGTA
AGTGAGAGTGTGAGTGGGTATTAGGTCTGTGAGTGGGTGGTGAGTGAGTGAGTAAAGAGTGG
GTTGGTGAGAAAGTGAGTGAGTCACTTGGTGGGTGCGTTAGTGGAAGCGTGATTGAGTGAGTGGA
GGTAGGTGAGTGAGTGAATTGGTGGGGGGGTGAGTGAGGTTAACGCTGTTCTGCTGTTCAATCACA
CCACATGTTGCCAGCTTACTGTGCAGGACGAATCCAGGGTTGTGTTAAATTTTATATGTTTATATA
TAACGATGGACGTGTCTGGATGTGGCGAATGTGTCAAGAGAATTATGCGGCTTTGTGCTGCTCCGC
GTATTTATTGCACGCGCGTTGGTACGCGGTTGATAAAGTAGTTCAAACATTTCCCAGCCATCTTT
GTCTGTTGTGAAAACCTACTCCAGGACCATCCATTTCAATATGTGTCTGCGTTCATGGAGTTATAC
ATGTTAAACTGTAGAGCGCAGATGAGCACACTTGAGCATTTCTTCAGTAAATCAGAATGTGTATAT
TTCAAAATTTACCAAATGCAATATCATCAAGCAAATTATGCAGCTCTATAGTAACATCGGAGTCAA
TGGTCCAGTGTGCCCTCGGCTGCCATTCCGACCTCCCTGGCCAGAATACACCCCGGTGAGGATCAG
TTATCCGTCAGAAGGCACGGTGCGGAATGAAAACATAAACACATAGTCGCTTAGTAGTATGCTGAT
TTAGGCACGCAAAATCCGAATGTGAATTACTGTGAATTGCATTACCTGTTACAG

Fig. 6c

DOMAIN 2B

AGGCCCCAGCTCCCTCCTCGGATGCTCACCTCGCCGTCAGGAAGGATATCAACCATCTGACACGCG
AGGAGGTGTACGAGCTGCGCAGAGCTATGGAGAGATTCCAGGCCGACACATCCGTTGATGGGTACC
AGGCTACGGTTGAGTATCACGGCTTACCTGCTCGATGTCCATTCCTCCGAGGCCACAAATAGGTTCG
CCTGTTGCATCCACGGCATGGCGACATTCCCTCATTGGCACAGACTGTTCTGTTACCCAGGTGGAAG
ATGCACTGATCAGGCGAGGATCCCCATAGGGGTCCCCCTACTGGGACTGGACTCAGCCTATGGCAC
ATCTCCCAGGACTTGCAGACAACGCCACCTATAGAGATCCCATCAGCGGAGACAGCAGACACAACC
CGTTCCACGATGTTGAAGTTGCCTTTGAAAATGGGCGTACAGAACGTCACCCAGATAGTAGATTGT
TTGAACAACCTCTATTTGGCAAACATACGCGTCTCTTCGACAGTATAGTCTATGCTTTTGAGCAGG
AGGACTTCTGCGATTTTGAAGTTCAATTTGAGATGACCCATAATAATATTCACGCCTGGATTGGTG
GCGGCGGGAAGTATTCCATGTCTTCTCTACACTACACAGCCTTCGACCCTATCTCCTACCTTCATC
ACTCCAACACTGACCGTCTCTGGGCAATTTGGCAAGCGTTGCAGATACGAAGAAACAAACCGTATA
AGGCTCATTGTGCTTGGTCTGAGGAACGCCAGCCTCTCAAACCTTTCGCCTTCAGTTCCCCACTGA
ACAACAACGAAAAAACCTACGAAACTCGGTGCCACCAACGTTTACGACTACGAAGGAGTCCTTG
GCTATACTTATGATGACCTCAACTTCGGGGGCATGGACCTGGGTGAGCTTGAGGAATACATCCAGA
GGCAGAGACAGAGAGACAGGACCTTTGCTGGCTTCTTTCTGTACATATTGGTACATCAGCGAATG
TTGAAATCATTATAGACCATGGGACTCTTCATACCTCCGTGGGCACGTTTGCTGTTCTTGCGGAG
AGAAGGAGATGAAATGGGGATTTGACCGTTTGTACAAATATGAGATTACAGATGAACTGAGGCAAC
TTAATCTCCGTGCTGATGATGGTTTCAGCATCTCTGTAAAGTAACTGATGTTGATGGCAGTGAGC
TGTCCTCTGAACTCATCCCATCTGCTGCTATCATCTTCGAACGAAGCCATA

INTRON 2B/2C (SEQ ID NO:129)

GTAAGTAGCTACCTGTTTATTCAATTTTTTCGCTTTGCCAATCAATTCATTTCAGCTTGAAATTCAA
TAATTGTGTTTTGCATGGCTGAAAACCAATTTGAACTCTTTTCTTTTCTCAGGTGCAACTCAAATA
AATAATCACTAATTGTTATGCACGCGGGTAGGGCATACTATATCCACATCGGTTCATCTCAA
ATGCAACAAATTGTCTTATTTCCGTTGGGACAAGCAAACCCCTTTCCTGTAATCTTGCCTTTGG
CATCCACTGGAATTAATGTTGACTGGTAATTGATACTGGCTCTCTTCTTGATAGAGTTAATATCT
ATAGTTTGTAATCTTTATGATTTTGCTATTTATATTTTCGACAGCATGCTATAGACACCCTAGACT
ATTGTATAGCCACTTGATTTGTTTTCCATTTATTATTTATAACAGAACATGGCTTGTAATTTTAA
TTTACCTTCCAG

DOMAIN 2C

TTGACCATCAGGACCCTCATCAGGACACAATCATCAGGAAAAATGTTGATAATCTTACACCCGAGG
AAATTAATTCTCTGAGGAGGGCAATGGCAGACCTTCAATCAGACAAAACCGCCGGTGGATTCCAGC
AAATTGCTGCTTTTCACGGGGAACCCAAATGGTGCCCAAGTCCCGATGCTGAGAAGAAGTTCTCCT
GCTGTGTCCATGGAATGGCTGTCTTCCCTCACTGGCACAGACTCCTGACCGTGCAAGGCGAGAATG
CCCTGAGAAAGCATGGATGTCTCGGAGCTCTCCCCTACTGGGACTGGACTCGGCCCTGTCTCACC
TACCTGATTTGGTAAGTCAGCAGAACTACACCGATGCCATATCCACCGTGGAAGCCCGAAACCCCT
GGTACAGCGGCCATATTGATACAGTTGGTGTGACACAACAAGAAGCGTCCGTCAAGAAGTGTATG
AAGCTCCCGGATTTGGTCATTATACTGGGGTCGCTAAGCAAGTGCTTCTGGCTTTGGAGCAGGATG
ACTTCTGTGATTTTGAAGTCCAGTTTGAGATAGCTCACAATTTTCATCCACGCTCTTGTGCGCGGAA
GCGAGCCATATGGTATGGCGTCACTCCGTTACACTACTTATGATCCAATTTTCTACCTCCATCATT
CTAACACTGACAGACTCTGGGCTATATGGCAGGCTCTACAAAAGTACAGGGGCAAACCTTACAATT
CCGCCAACTGTGCCATTGCTTCTATGAGAAAACCCCTACAGCCCTTTGGTCTGACTGATGAGATCA
ACCCGGATGATGAGACAAGACAGCATGCTGTTCCCTTTCAGTGTCTTTGATTACAAGAACAACCTTCA
ATTATGAATATGACACCCTTGACTTCAACGGACTATCAATCTCCAGCTGGACCGTGAAGTGTGAC
GGAGAAAGTCTCATGACAGAGTATTTGCCGGATTTTGTGCTGCATGGTATTTCAGCAGTCTGCACTAG

Fig. 6d

TTAAATTCTTTGTCTGCAAATCAGATGATGACTGTGACCACTATGCTGGTGAATTCTACATCCTTG
GTGATGAAGCTGAAATGCCATGGGGCTATGATCGTCTTTACAAATATGAGATCACTGAGCAGCTCA
ATGCCCTGGATCTACACATCGGAGATAGATTCTTCATCAGATACGAAGCGTTTGATCTTCATGGTA
CAAGTCTTGGAAGCAACATCTTCCCCAAACCTTCTGTCATACATGACGAAGGGGCAG

INTRON 2C/2D (SEQ ID NO:130)

GTGAGAACATTGATAATAGTTCAAATGAAGTATATCCGATTCAAGCTGTCGATACAAGATGAGATA
CATAATCACAATGTTTGTATTAGATATCTCTCTTAATTTAATGCCGCTTTTATCAATATTCGAGCA
ATCCTTCAGCAACATACACCAGCAAATGTTTCATCAACAGACTATATTATTTAATATTTTAAAAAT
CCTTCTCTGTTGTTATAAATACTTAAAGTATCGAATTCCTTGAATGCGTCTTCTCTGCAGCATATA
GTTAAGTTGTTGTGTTTCTCTGTCAG

DOMAIN 2D

GTCACCATCAGGCTGACGAGTACGACGAAGTTGTAAGTCTGCAAGCCACATCAGAAAGAATTTAA
AAGATCTGTCAAAGGGAGAAGTAGAGAGCCTAAGGTCTGCCTTCCTGCAACTTCAGAACGACGGAG
TCTATGAGAATATTGCCAAATTCACGGCAAGCCTGGGTTGTGTGATGATAACGGTTCGCAAGGTTG
CCTGTTGTGTCCATGGAATGCCACCTTCCCCCAGTGGCACAGACTCTATGTCTCCAGGTGGAGA
ATGCTTTGCTGGAGAGAGGATCTGCCGTCTCTGTGCCATACTGGGACTGGACTGAAACATTTACAG
AGCTGCCATCTTTGATTGCTGAGGCTACCTATTTCAATTCCCGTCAACAAACGTTTGACCCTAATC
CTTTCTTCAGAGGTAAAATCAGTTTTTGAGAATGCTGTTACAACACGTGATCCCCAGCCTGAGCTGT
ACGTTAACAGGTACTACTACCAAAACGTCATGTTGGCTTTTGAACAGGACAACACTACTGCGACTTCG
AGATACAGTTTGAGATGGTTCACAATGTTCTCCATGCTTGGCTTGGTGGAAAGAGCTACTTATTCTA
TTTCTTCTCTTGATTATTCTGCATTTCGACCCTGTGTTTTTCTTCCCATGCGAACACAGATAGAT
TGTGGGCCATCTGGCAGGAGCTGCAGAGGTACAGGAAGAAGCCATACAATGAAGCGGATTGTGCCA
TTAACCTAATGCGCAAACCTCTACATCCCTTCGACAACAGTGATCTCAATCATGATCCTGTAACCT
TTAAATACTCAAAACCCACTGATGGCTTTGACTACCAGAACAACCTTTGGATACAAGTATGACAACC
TTGAGTTCAATCATTTCAGTATTCCCAGGCTTGAAGAAATCATTCGTATTAGACAACGTCAAGATC
GTGTGTTTGCAGGATTCCCTCCTTCACAACATTGGGACATCCGCAACTGTTGAGATATTTCGTCTGTG
TCCCTACCACCAGCGGTGAGCAAACTGTGAAAACAAAGCCGGAACATTTGCCGTACTCGGAGGAG
AAACAGAGATGGCGTTTCATTTTGACAGACTCTACAGGTTTGACATCAGTGAAACACTGAGGGACC
TCGGCATAACAGCTGGACAGCCATGACTTTGACCTCAGCATCAAGATTCAAGGAGTAAATGGATCCT
ACCTTGATCCACACATCCTGCCAGAGCCATCCTTGATTTTTTGTGCCTGGTTCAA

INTRON 2D/2E (SEQ ID NO:131)

GTAAGAAAGTTTCACTGTCTAAATCTTTTTTTATGATAGAGGGTAGAGAAGTGGAGACAATGTGAC
AATATATTGAATAAAGTTGTTTAAAATTTATAACTCTCATAAGTTCATATTATGCTGAAGCTGTAG
CCATCTATAACTGTGTAACATGAAATGTTAAGACATTAACCTAAATACTTCAGCTGATAACAAAAC
AATGTTAATACATACGTCAATGTAACATTTTCTTATCTTTAGGTTATAGCATAAAACACTTCAGAGA
TACAGTGACGAAAACCTCTATTTAAATATTTTCAG

DOMAIN 2E

GTTCTTTCTGCGTCCTGATGGGCATTTCAGATGACATCCTTGTGAGAAAAGAAGTGAACAGCCTGA
CAACCAGGGAGACTGCATCTCTGATCCATGCTCTGAAAAGTATGCAGGAAGACCATTACCTGATG
GGTTCCAAGCCATTGCCTCTTTCCATGCCCTGCCACCACTCTGCCCTTCACCATCTGCAACTCACC
GTTATGCTTGCTGTGTCCACGGCATGGCTACATTTCCCCAGTGGCACAGACTGTACACTGTACAGT

Fig. 6e

TCCAGGATGCACTGAGGAGACATGGAGCTGCAGTAGGTGTACCGTATTGGGATTGGCTGCGACCGC
AGTCTCACCTACCAGAGCTTGTCCACCATGGAGACATACCATGATATTTGGAGTAACAGAGATTTCC
CCAATCCTTTCTACCAAGCCAATATTGAGTTTGAAGGAGAAAACATTACAACAGAGAGAGAAGTCA
TTGCAGACAAACTTTTTGTCAAAGGTGGACACGTTTTTTGATAACTGGTTCTTCAAACAAGCCATCC
TAGCGCTTGAGCAGGAAACTACTGTGACTTTGAGATTAGTTTGAATTCCTTCAACAGGCGTTC
ACACGTGGGTGCGAGGCAGTCGTACCCACTCTATCGGACATCTCCATTACGCATCCTACGACCCTC
TTTTCTACCTCCACCATTCCCAGACAGACCGTATTTGGGCAATCTGGCAAGAACTCCAGGAACAGA
GAGGGCTCTCAGGTGATGAGGCTCACTGTGCTCTCGAGCAAATGAGAGAACCATTGAAGCCTTTCA
GCTTCGCGCTCCTTATAACTTGAATCAGCTAACACAGGATTTCTCCCGACCCGAGGACACCTTCG
ACTACAGGAAGTTTGGTTATGAATATGACAATTTAGAATTCCTAGGAATGTCAGTTGCTGAACTGG
ATCAATACATTATTGAACATCAAGAAAATGATAGAGTATTCGCTGGGTTCTGTTGAGTGGATTTCG
GAGGTTCCGCATCAGTTAATTTCCAGGTTTGTAGAGCTGATTCACATGTCAGGATGCTGGGTACT
TCACCGTTCTTGGTGGCAGTGCTGAGATGGCGTGGGCATTTGACAGGCTATACAAATATGACATTA
CTGAAACTCTGGAGAAAATGCACCTTCGATATGATGATGACTTCACAATCTCTGTCTGAGTCTGACCG
CCAACAACGGAAGTGTCTGAGCAGCAGTCTAATCCCAACACCGAGTGTATATTCAGCGGGGAC
ATC

INTRON 2E/2F-1 (SEQ ID NO:132)

GTAAGTAGTAACTGCTCAGATTGTTTTTCATAATTACTCCACTATTAAGTAAAAAGTACTAGTAAT
TCAATAGTACTGTTTCACAGAGAAATGTAACACAATAGACCACAGAGTCCATTTGTTAAACGCCTTT
GGCTTGGTAAAGTCTGAGATTTTGGTGACTGATGGAAAGCTAAAATATATTTTGACAG

DOMAIN 2F-1 (1st part of domain f)

GTGACATAAATACCAAGAGCATGTCAGCGAACCGTGTTCCCGTGAGCTGAGCGATCTGTCTGCGA
GGGACCCGCTAGTCTCAAGTCTGCTCTGCGAGACCTACAGGAGGATGATGGCCCCAACGGATACC
AGGCTCTTGACGCCTTCCATGGGCTACCAGCAGGCTGCCATGATAGCCAGGGAAATGAG

INTRON 2F-1/2F-2 (SEQ ID NO:133)

GTATATTTAAGTATTTTATCTTACGCATGACCCTGACCCTATTTATTTTTTTTTTAATCCTCGGATT
TGTTTAAATCCTGTTACCAGCGAAGGTCCGGGTAGAATTGATCTTCAGTCAACTATTCTTGTCTGTA
GGACTAACGAGTTGTCTGGCTTGCTTACTCGGTTGACACGTGTCAACGGATCCCAATTGCAATTAG
ATCGATGCTCATGCTGTTGATCCCTGGATTGCCTGGTCCGGACTCCACATACCGCCGCATATTGC
TGGTATATTGTCGAATGCGACGCTAAACAGCAAGCCAACCAACAATACTGAGACCTGGTGGTACAT
GTCAGTTCTCTATTGCTGGGGTTCCAAACATAGCCATCAGTTGAAATATTTTCATACATAGAAGAAT
ACCTCTGAATATGATGATGAAACATTTACTTAGACTTGCTGTGAGCCCCAGGCAAAATGCACTGT
AAAAATACACTGACAGAGGATTAGGCATTCTTGGGAGTACTGTATAGTTAGTTGCATACATATTAG
CGTTCCCTCACTAAAACGAATCTCTGAATGCTATCAATTAAAGATCATGATGCTTTGATTGTGTCT
ACTGTATTTAAATGGTGTAAAGATTTGCAATTACAATATACACAAACACGTTTCCTGCATCTCGG
AGAATGCAATCTTTCGTTGTACGCGTCTGTTTTTCATATTTTATGCATGTAGTTGCACTACTTAG
CGTCCAATAAATCCATTACAAAATCACACAAACAAACGATTTTAGGAATGTGACTGTAGCTGCAA
CGAATATACCTGATCCTTTCTTGTTCAG

DOMAIN 2F-2 (2nd part of domain f)

ATCGCATGTTGCATTCACGGTATGCCGACCTTCCCCAGTGGCACAGACTGTACACCCTGCAGTTG
GAGATGGCTCTGAGGAGACATGGATCATCTGTGCGCATCCCCTACTGGGACTGGACAAAGCCTATC

Fig. 6f

TCCGAAC TCCCCTCGCTCTTCACCAGCCCTGAGTATTATGACCCATGGCATGATGCTGTGGTAAAC
AACCCATTCTCCAAAGGTTTTGTCAAATTTGCAAATACCTACACAGTAAGAGACCCACAGGAGATG
CTGTTCCAGCTTTGTGAACATGGAGAGTCAATCCTCTATGAGCAAACCTCTTCTTGCTCTAGAGCAA
ACCGACTACTGTGATTTTGAGGTACAGTTTGAGGTCCCTCCATAACGTGATCCACTACCTTGTGGC
GGACGTCAGACCTACGCATTGTCTTCTCTGCATTATGCATCCTACGACCCATTCTTCTTTATACAC
CATTCTTTGTGGATAAGATGTGGGTAGTATGGCAAGCTCTTCAAAGAGGAGGAACTTCCATAC
AAGCGAGCTGACTGTGCTGTCAACCTAATGACTAAACCAATGAGGCCATTTGACTCCGATATGAAT
CAGAACCCATTCACAAAGATGCACGCAGTTCCCAACACACTCTATGACTACGAGACACTGTACTAC
AGCTACGATAATCTCGAAATAGGTGGCAGGAATCTCGACCAGCTTCAGGCTGAAATTGACAGAAGC
AGAAGCCACGATCGCGTTTTTGTCTGGATTCTTGCTTCGTGGAATCGGAACCTTCTGCTGATGTCAGG
TTTTGGATTTGTAGAAATGAAATGACTGCCACAGGGGTGGAATAATTTTCATCTTAGGTGGAGCC
AAGGAAATGCCATGGTCATTTGACAGAACTTCAAGTTTGATATCACCCTGTACTCGAGAAAGCT
GGCATTAGCCAGAGGACGTGTTTGATGCTGAGGAGCCATTTTATATCAAGGTTGAGATCCATGCT
GTTAAACAAGACCATGATACCATCGTCTGTGATCCCAGCCCCAACTATCATCTATTCTCTGGGGAA
G

INTRON 2F-2/2G-1 (SEQ ID NO:134)

GTGAGAGAACCAGTAATAGCTACTGTCTACAAAGAATGTGTTTCAATTTAAAGACCTGACTGTAGGCC
GATGGCTGCTGTCATCTCCTCCGCCTCCTCCTCCTGTTCCCTCCTCCGAAGGGGTCAGCTTCAGGTT
CTCTTGCCAATATGCCAAGCAGACCTCCTGAGCAGGCAGTATATATACGTAAGGGAAGCAAGTATG
GACCATCGCGCGGCATGTAGAGATACAATGATCAGCTGTCTGCTGTTCCACTCCTGTCAGACAATG
AGATAAACATGAATACAGTATTACTCAGCAGCGTTCCAATTTTCAACCCTCGTATTTATTAATAA
AGGAATTTTAAATATATTTTTCTCCTTGTTGAAATATTTTAGTAACTGTTAATCGATATAGAGTGG
AGTAGTGACGCTTTATTTTCGGTTCATTCTCGAAACAAAAATATAATAGTCCACTGAACTCTCTTAA
ATTGTTTTTACAACCTTCAACTGCCACAGACGTAATCCCTCACGTTATTTTGAGCTGACAACGTGT
TGAATTGAGTGTGTTCCGAATTCTAAATAAGCATGTATATATTTACGTCTCATGCAAGTAATATAT
GTTTAACTGATGACGTCACCTGGTGACCACTGATTTAGTTTCTTTGTCATAATTGCAGTTTCTGTT
GTCACGGGGACGGTGGGGAAGCCAGGTTCTCCTGTCACGCTGAATATCCCGTTCGAATCCCCCAC
ATGGGTACAAAGTGTGATGCCTATTTCTGGTGTCCCCACCGTGATATTGCTGGAATAAGTGGCTT
AATACCATATACACTCACTCTATTGTCACTACTGCCACCGGCTCACACCTCTGATGCTTCTGTT
CTATCCAG

DOMAIN 2G-1 (1st part of domain g)

GTCGCGCTGCTGACAGTGCACACTCAGCCAACATTGCTGGCTCTGGGGTGAGGAAGGACGTCACGA
CCCTCACTGTGTCTGAGACCGAGAACCTAAGACAGGCTCTTCAAGGTGTCATCGATGATACTGGTC
CCAATGGTTACCAAGCAATAGCATCCTTCCACGGAAGTCCTCCAATGTGCGAGATGAACGGCCGCA
AGGTTGCCTGTTGTGCTCACG

INTRON 2G-1/2G-2 (SEQ ID NO:135)

GTAATTAATGGATGTGAAGTCAATGTCCGAGGGTATAATAAGGATTTAAATACTTCAGTCGTGTAA
TACTGTATGACATGTGTATTGGATGGTGTAGGTATTACAGGTATAAGGCCAGTGTGTGTTGGGAC
GGTTACTTTTCTGCACTAGTAATAAGCATTGTATTTAGCTAGCTTTTATCATATAACTTTAGTTTC
ATGGTTTGTGGCAATTGAAATCGAAATTTTCTTTCATTTCAAGGTTATCGCACTCGTGTGTTAGAA
TAGTTACTATGCTGCATTGAGAATAACACTATAGTAATAAAGCATATCATACAGTAAGAATAACAC
TATAGTAATAAAGTATATCATACAGTAAGAATGTCATTGTATGATAAATAGGTTATCACACTCGTG
TGTTTTAGAATGGTTACTATCCCAGGAATAACCACTATGTATTACATGTATATTGGGCAGTGTAAG
TAGTAGCATTGTATATTAAATCAGTATATCGTGCTTCAAACACCAGGATATATGGGGTATACAGT

Fig. 6g

GGGCAGTGTAAGTAGCAACATTGTATATTAAATCAGTATATCGTACTTCAAAACACCAGGATTATG
GGGTATACAGTGGGCAGTGTAAGTAGTAGCATTGTATATTAAATCAGTATATCGTACTTCAAAACA
CCAGGATATAATTTCAGTATATCGTGCTTCAAAACACCAGGATATAATTTCAGTATATCGTGCTTCAA
AACACCAGGATATATGGGATATACAGTGCGGGTTTGCATACAACCTCCACCCTTTACAG

DOMAIN 2G-2 (2nd part of domain g)

GTATGGCCTCCTTCCCACACTGGCACAGACTGTATGTGAAGCAGATGGAAGACGCCCTGGCTGACC
ACGGATCACATATCGGCATCCCTTACTGGGACTGGACAACCTGCCTTCACAGAGTTACCCGCCCTTG
TCACAGACTCCGAGAACAATCCCTTCCATGAG

INTRON 2G-2/2G-3 (SEQ ID NO:136)

GTCAGTTTAGTCTCCTGTCTGAGCTAACGATACCAATTTCTTATTTTCGAGAACCACGATGACGAG
AAAACAAGCAATATAGATATAGATGCAGTATAGATCAAGTTAATGAATTCATTGCTATATGTTTGC
TTGTAATAAACTTTAAGAAAACGAGAGCATGCACACAAATGAAACAAACAATTATGTGTTTGATAG
GAATATGATATATGTATTTGGGGGCTGACGTGAGCAGGGTTGAAGGGACAGTTTACATTGTGAGTA
ACACTGGGAGTATTCTTTGATCCACAATATATAGTTTCATTGTGTTTCAGCAGTTACAACCTAACATT
ATATCATACATTACGTCGTAACATGCTTCTTTTGTCTCTTCTGCCAG

DOMAIN G-3 (3rd part of domain g)

GGTCGCATTGATCATCTCGGTGTAACCACGTCACGTTCCCCCAGAGACATGCTGTTTAAACGACCCA
GAGCAAGGATCAGAGTCGTTCTTCTATAGACAAGTCCTCCTGGCTTTGGAGCAGACTGACTACTGC
CAGTTTCGAAGTCCAGTTTGAGCTGACCCACAACGCCATTCACTCCTGGACAGGTGGACGTAGCCCT
TACGGAATGTCGACCCTCGAGTTCACAGCCTACGATCCTCTCTTCTGGCTTACCACTCCAACACC
GACAGAATCTGGGCTGTCTGGCAAGCACTGCAGAAATACCGAGGACTCCCATACAACGAAGCACAC
TGTGAAATCCAGGTTCTGAAACAGCCCTTGAGGCCATTCAACGATGACATCAACCACAATCCAATC
ACCAAGACTAATGCCAGGCCTATCGATTCAATTTGATTATGAGAGGTTTAACTATCAGTATGACACC
CTTAGCTTCCATGGTAAGAGCATCCCTGAACTGAATGACCTGCTCGAGGAAAGAAAAAGAGAAGAG
AGAACATTTGCTGCCTTCTTCTCGTGGAATCGGTTGCAGTGCTGATGTCGTCTTTGACATCTGC
CGCCCCAATGGTGACTGTGTCTTTGCAGGAACCTTTGCTGTGCTGGGAGGGGAGCTAGAAATGCCT
TGGTCCTTCGACAGACTGTTCCGCTATGACATCACCAGAGTCATGAATCAGCTCCATCTCCAGTAT
GATTTCAGATTTTCAGTTTCAGGGTGAAGCTTGTTGCAACCAATGGCACTGAGCTTTCATCAGACCTC
CTCAAGTCACCAACAATTGAACATGAACTTGAG

INTRON 2G-3/2H (SEQ ID NO:137)

GTATGTTATCTTATTATCAAATGTGTAATCAGATACTGGAGACGTTTTTCATATTAACCTGGTCAGC
ATTAGTTGATGATTTTGGTGCGATATTGACGACAAGGAGTTAAGCATTAACACGTTCAACACATCT
TTAATCTGATATGAGAAGGGAATAAATTGATCCAGTATTGATGATTGAAGTTAGATTAACAGTGAA
AGATATACCAGTTTTGATAATCGTATAAAACAGTAGCAGAATTGTATCGTGAAAATAAATGTGGG
AAGGCGAACGCCAAGCAGATTTTAGATTACGATCGTGTGCTAGAATAATTACAATAACCCAGACG
TCGGAATGTGGTTGTCTATGGCAATAGTTACGATTAATTGCTAACATGCACGATTTACCTATTTCC
AG

DOMAIN 2H

CCCACAGAGGACCAGTTGAAGAAACAGAAGTCACTCACCAAATACTGACGGCAATGCACACTTCC
ATCGTAAGGAAGTTGATTCGCTGTCCCTGGATGAAGCAAACAACCTGAAGAATGCCCTTTACAAGC

Fig. 6h

TACAGAACGACCACAGTCTAACAGGATACGAAGCAATCTCTGGTTACCATGGATACCCGAATCTGT
GTCCGGAAGAAGGCGATGACAAATACCCCTGCTGCGTCCACGGAATGGCCATCTTCCCCCACTGGC
ACAGACTCTTGACCATCCAACCTGGAAAGAGCTCTCGAGCACAATGGTGCACCTGCTTGGTGTTCCTT
ACTGGGACTGGACCAAGGACCTGTCGTCACCTGCCGGCGTTCTTCTCCGACTCCAGCAACAACAATC
CCTACTTCAAGTACCACATCGCAGGTGTTGGTCACGACACCGTCAGAGAGCCAACTAGTCTTATAT
ATAACCAGCCCCAAATCCATGGTTATGATTATCTCTATTACCTAGCATTGACCACGCTTGAAGAAA
ACAATTACTGTGACTTTGAGGTTTCAAGTATGAGATCCTCCACAACGCCGTCCACTCCTGGCTTGGAG
GATCCCAGAAGTATTCCATGTCTACCCTGGAGTATTCGGCCTTTGACCCTGTCTTTATGATCCTTC
ACTCGGGTCTAGACAGACTTTGGATCATCTGGCAAGAACTTCAGAAGATCAGGAGAAAGCCCTACA
ACTTCGCTAAATGTGCTTATCATATGATGGAAGAGCCACTGGCGCCCTTCAGCTATCCATCTATCA
ACCAGGACGAGTTTACCCGTGCCAACTCCAAGCCTTCTACAGTTTTTGGACAGCCATAAGTTCGGCT
ACCATTACGATAACCTGAATGTTAGAGGTCACAGCATCCAAGAAGTCAACACAATCATCAATGACT
TGAGAAACACAGACAGAATCTACGCAGGATTTGTTTTGTTCAGGCATCGGTACGTCTGCTAGTGTC
AGATCTATCTCCGAACAGATGACAATGACGAAGAAGTTGGAACCTTCACTGTCTGGGAGGAGAGA
GGGAAATGCCATGGGCCTACGAGCGAGTTTTCAAGTATGACATCACAGAGGTTGCAGATAGACTTA
AACTAAGTTATGGGGACACCTTTAACTTCCGACTAGAGATCACATCCTACGATGGATCGGTGGTAA
ACAAGAGCCTACCCAATCCTTTTCATCATCTACAGACCTGCCAATCATGACTACGATGTTCTTGTTA
TCCCAGTAGGAAGAAACCTTCACATCCCTCCCAAAGTTGTCGTCAAGAGAGGCACCCGCATCGAGT
TCCACCCAGTCGATGATTACGTTACGAGACCAGTTGTTGATCTTGGAAGCTACACTGCACTCTTCA
ACTGTGTGGTACCACCGTTTACATACCGCGGATTCGAAGTGAACCACGTCTATTCTGTCAAGCCTG
GTGACTACTATGTTACCGGACCAACGAGAGACCTTTGCCAGAATGCAGATGTCAGGATTCATATCC
ATGTTGAGGATGAGTAA

3'UTR

CGCAACAG

INTRON 3'UTR (SEQ ID NO:138)

GTGAGATAAGAAACCCTTCTAACAGTAATACGACACCACATTACAGCTTAAACATGATTGCCATCG
ATGTTTTTCATGTGTAGTATACGCTTTTTCAGTTCTACATAATTTTGTTTTTTCAAATCAAGTTTAGCA
AATGAATCTATCACTGGAAAATAGGGTAGGGTAGCCAAGTGGTTAAAGCGGTCACTGATCACGCCA
AAGACGAGTGTCTAACCTGCATGGGTACAAAAGTGAAGACCATTGCTGGTGTCTACCGCCGTAAT
ATTGTTTTTTAGTATTGCTAAAACTTATACTCACCCTATGCGCTGTAAAAGTGGAATAATAATCATAT
TTCAACAAAAGCACAAAACCATTTTCATTTTCATGAAAGCCTCTTGTTACCTGAAAGACGCAAGAG
AACAATAGTTCCTAACATTATTTTCAGACATTGGAAATGTCCTGCACGTGTAAACCATATATCCTT
TGAAATTTTTTACGACTGCATCGTATACAATTTATGATATAAAATTTAAAACTTTATTTTCAG

3'UTR

GTTTCTTGGTCTCCACATATTCACACATCAGCACCAAACGGTTTTCGAAGGACATTGGCGTTCTTCT
CTGGCAATGCATTTCAATACAACATTGAAAATGACTTCAGCATATCAGTGTGCTTCGAACGTGTTC
CGGAAGTACTCAAATGTGCTATGACTGAATTATTGTACATACATACTTATTGATGTTCAATAAAT
AAATGTTGAAACG

Fig. 7a**Primary structure of the Hth2 protein****DOMAIN A (SEQ ID NO:156)**

GLPYWDWTQHLLTQLPDLVSDPLFVDPEGGKAHDNAWYRGNIKFENKKTARAVDDRLFEEKVGPGEN
RLFEGILDALEQDEFNCNFEIQFELAHNAIHLYLVGGRHTYSMSHLEYTSYDPLFFLHHSNTDRIFAI
WQRLQVLRGKDPNTADCAHNLIHEPMEPFRRDSNPLDLTRENSKPIDSFYAHGLGYQYDDLTLNGM
TPEELNSYLHERSGKEGVFASFRLSGFGGSANVVVYACRPAHDEMAVDQCDKAGDFFVLGGPTMP
WRFYRAFHFDVTDSDIDNIDKDRHGHYYVKAELFSVNGSALPNDLLPQPTISHRPARGHVDEAPAPS
SDAHLAVRKDINHLLTREEVYELRRAMERFQADTSVDGYQATVEYHGLPARCPFPEATNRFACCIHG
MATFPHW

DOMAIN B

HRLFVTQVEDALIRRGSPIGVPYWDWTQPM AHL PGLADNATYRDPISGDSRHNPFHDVEVAFENGR
TERHPDSRLFEQPLFGKHTRLFDSIVYA FEQEDFCDFEVQFEMTHNNI HAWIGGGGKYSMSSLHYT
AFDPISYLHHSNTDRLWAIWQALQIRRNKPYKAHCAWSEERQPLKPFASFSSPLNNNEKTYENSVPT
NVYDYEGLVGYTYDDL NFGGMDLGQLEEYIQRQRQDRTFAGFFLSHIGTSANVEI IIDHGT LHTS
VGTF AVLGG EKEMK WGFDRLYKYEITDEL RQLNLRADDGFSISVKVTDVDGSELSS ELIPSAAIIF
ERSH

DOMAIN C

IDHQDPHQDTIIRKNVDNLTPEEINSLRRAMADLQSDKTAGGFQQIAAFHGEPKWCPSPDAEKKFS
CCVHGM AVFPHWHRLTVQGENALRKHGCLGALPYWDWTRPLSHLPDLVSQQNYTDAISTVEARNP
WYSGHIDTVGVDTTTSVRQELYEAPGFGHYTGVAQVLLALEQDDFCDFEVQFEIAHNFIHALVGG
SEPYGMASLRYTTYDPIFYLHHSNTDRLWAIWQALQKYRGKPYNSANCAIASMRKPLQPFGLTDEI
NPDDETRQH AVPF SVFDYKNNFN EYDTLDFNGLSISQLDRELSRRKSHDRVFAGFLLHGIQQSAL
VKFFVCKSDDDCDH YAGEFYILGDEAEMPWG YDRLYKYEITEQLNALDLHIGDRFFIRYEA FDLHG
TSLGSNIFPKPSVIHDEGA

DOMAIN D

GHHQADEYDEVVTAASHIRKNLKDLSKGEVESLRSAFLQLQNDGVYENIAKFHGKPGLCDDNGRKV
ACCVHGMPTFPQWHRLYVLQVENALLERGS AVSVPYWDWTETFTELPSLIAEATYFNSRQQT FDPN
PFRGKISFENAVTTRDPQPELYVNRYYYQNVMLAFEQDNYCDFEIQFEMVHNVLHAWLGGRATYS
ISSLDYSAFDPVFFLHHANTDRLWAIWQELQRYRKKPYNEADCAINLMRKPLHPFDNSDLNHDPVT
FKYSKPTDGFQYQNNFGYKYDNLEFNHFSIPRLEEII RIRQRQDRVFAGFLLHNIGTSATVEIFVC
VPTTSGEQNCENKAGTF AVLGG ETEMAFHFDRLYRFDISETLRDLGIQLD SHDFDLSIKIQGVNGS
YLDPHILPEPSLIFVPGSS

DOMAIN E

SFLRPDGHSSDDILVRKEVNSLTRETASLIHALKSMQEDHSPDGFQAIASFHALPPLCPSPSATHR
YACCVHGMATFPQWHRLYTVQFQDALRRHGAAVGVPYWDWLRPQSHLP ELVTMETYHDIWSNRDFP
NPFYQANIEFEGENIT TEREVIADKLFVKGGHVFDNWFFKQAILALEQENYCDFEIQFEILHNGVH
TWVGGSRTHSIGHLHYASYDPLFY LHHSQTDRIWAIWQELQEQRGLSGDEAHCALEQMREPLK PFS
FGAPYNLNQLTQDFSRPEDTFDYRKFGYEYDNLEFLGMSVAELDQYII EHQENDRVFAGFLLSGFG
GSASVNFQVCRADSTCQDAGYFTVLGGSAEMAWAFDRLYKYDITETLEKMHLRYDDDF TISVSLTA
NNGTVLSSSLIPTPSVIFQRGH

Fig. 7bDOMAIN F

RDINTKSMSANRVRRELSDSL SARDPSSLKSALRDLQEDDGPNGYQALAAFHGLPAGCHDSQGNEIA
CCIHGMPTFPQWHRLYTLQLEMALRRHGSSVAIPYWDWTKPISELPSLFTSPEYYDPWHDVNNP
FSKGFVKFANTYTVRDPQEMLFQLCEHGESI LYEQTLLALEQTDYCDFEVQFEVLHNVIHYLVGGR
QTYALSSLHYASYDPFFFIHHSFVDKMVWVWQALQKRRKLPYKRADCAVNLMTKPMRPFDSMDNQN
PFTKMHAVPNTLYDYETLYSYDNLEIGGRNLDQLQAEIDRSRSHDRVFAGFLLRGIGTSADVRFW
ICRNENDCHRGGIIFILGGAKEMPWSFDRNFKFDITHVLEKAGISPEDVFDAEEPFIKVEIHAVN
KTMIPSSVIPAPTI IYSPGE

DOMAIN G

GRAADSAHSANIAGSGVRKDVTTTLTVSETENLRQALQGVIDDTGPNGYQAIASFHGSPPMCEMNGR
KVACCAHGMASFPHWRLYVKQMEDALADHGSHIGIPYWDWTTAFTELPALVTDSENNPFHEGRID
HLGVTTSRSPRDMLENDPEQGSSEFFYRQVLLALEQTDYCFEVQFELTHNAIHSWTGGRSPYGM
TLEFTAYDPLFWLHHSNTDRIWAVWQALQKYRGLPYNEAHCEIQVLKQPLRPFNDNDINHNPI TKT
NARPIDSFDYERFNYQYDTLSFHGKSIPELNDLLEERKREERTFAAFLLRGIGCSADVDFDICRPN
GDCVFAGTFAVLGGELEMPWSFDRLEFRYDITRVMNQLHLQYDSDFSFRVKLVATNGTELSDDLKSP
TIEHEL

DOMAIN H

GAHRGPVEETEVTHTQNTDGNNAHFHRKEVDSLDEANNLKNALYKLQNDHSLTGYEAI SGYHGYPN
LCPEEGDDKYPCCVHGMAIFPHWRLLT IQLERALEHNGALLGV PYWDWTKDLSSLPAFFSDSSNN
NPYFKYHIAGVGHD TVREPTSLIYNQPQIHGYDYLYYLA LTTLEENNYCDFEVQYEILHNAVH SWL
GGSQKYSMSTLEYSAFDPVFMILHSGLDRLWI IWQELQKIRRKPYNFAK CAYHMMEEPLAPFSYPS
INQDEFTRAN SKPSTVFD SHKFGYHYDNLNVRGHSIQELNTI INDLRNTDRIYAGFVLSGIGTSAS
VKIYLRTDDNDEEVGTFTVLGGEREMPWAYERVFKYDITEVADRLKLSYGDTFNFRL EITSYDGSV
VNKSLPNPFI IYR PANHDYDVLVI PVGRNLHI PPKVVVKRGTRIEFHPVDDSVTRPVVDLGSYTAL
FNCVVP PFTYRGFELNHVYSVKPGDYVTGPTRDLCQNADVRIHIHVEDE

Fig 8a**Genomic sequence of the KLH1 gene****DOMAIN 1B**

GGCCTACCGTACTGGGACTGGACTGAACCCATGACACACATTCCGGGTCTGGCAGGAAACAAAACCT
TATGTGGATTCTCATGGTGCATCCACACAAATCCTTTTCATAGTTCAGTGATTGCATTTGAAGAA
AATGCTCCCCACACCAAAAGACAAATAGATCAAAGACTCTTTAAACCCGCTACCTTTGGACACCAC
ACAGACCTGTTCAACCAGATTTTGTATGCCTTTGAACAAGAAGATTACTGTGACTTTGAAGTCCAA
TTTGAGATTACCCATAACACGATTCACGCTTGGACAGGAGGAAGCGAACATTTCTCAATGTCGTCC
CTACATTACACAGCTTTCGATCCTTTGTTTTACTTTTACCATTCTAACGTTGATCGTCTTTGGGCC
GTTTGGCAAGCCTTACAGATGAGACGGCATAAACCCCTACAGGGCCCCACTGCGCCATATCTCTGGAA
CATATGCATCTGAAACCATTTCGCTTTTTCATCTCCCCTTAACAATAACGAAAAGACTCATGCCAAT
GCCATGCCAAACAAGATCTACGACTATGAAAATGTCTCCATTACACATACGAAGATTTAACATTT
GGAGGCATCTCTCTGGAAAACATAGAAAAGATGATCCACGAAAACCAGCAAGAAGACAGAATATAT
GCCGGTTTTCTCCTGGCTGGCATACTGTTTACGCAAATGTTGATATCTTCATTAATAACTACCGAT
TCCGTGCAACATAAGGCTGGAACATTTGCAGTGCTCGGTGGAAGCAAGGAAATGAAGTGGGGATTT
GATCGCGTTTTCAAGTTTGACATCACGCACGTTTTTGAAGATCTCGATCTCACTGCTGATGGCGAT
TTCGAAGTTACTGTTGACATCACTGAAGTCGATGGAATAAAGTTCATCCAGTCTTATTCACAT
GCTTCTGTCATTCGTGAGCATGCACGTGGTAAGCTGAATAGAG

INTRON 1B/1C (SEQ ID NO:139)

GTTTTGTAATAATTATGTAGAATTCTTTACCTCAGAATAAGATGAGGTCACATGGGTTTTGCAAAA
CTATTACGTTTCGAATTAATATTAATAATACCGGACCTCCACTGGTACATATTTATCTTTATAACG
ATAATAGCGATGATGATGATGATGATGATGATGATGATGATGATGATAATGATGATGCCGGTATTG
CACGTAATCCAGCCGACTTAGATGACACCCTAAGGGTGCAGAAAGTATAACAATTAGATTGCGTTT
GCATCTGTGTATGCGTGTGCTTTAACCAAAAGTCAAATAAAAAGTGCAAACCTTAGTTTTATTCAT
TTGATAGAGCCTTTTACGATAAGAACAATGTAATAAATTAGAACAATACTGAAACCTCCGAAAGAA
GGCCTGTTTGTCAAGAGAGGTATCGACATGATTGACTTATAAACCTGTGCTTCTATATTTTGAAC
TGTCCACTTTCTTGTGTGTGTACTGTAATCACATCGCACTATGGCTGCAAGACGTGTACGAGTAC
ACTATATACTTACCTAATGACCAACCACAAGGCTGGCTTTGTTAATATTGTTATTTTACAGAAATA
AACACAGAATTCCAGCATTTGGCTGGTGTATTTAGCAAAACACCGATATGACACTCATGTTTTATT
ACATTTTTTTTCAG

DOMAIN 1C

TTAAATTTGACAAAGTGCCAAGGAGTCGTCTTATTCGAAAAAATGTAGACCGTTTGAGCCCCGAGG
AGATGAATGAACTTCGTAAAGCCCTAGCCTTACTGAAAGAGGACAAAAGTGCCGGTGGATTTTCAGC
AGCTTGGTGCATTCCATGGGGAGCCAAAATGGTGTCTTAGTCCCGAAGCATCTAAAAAATTTGCCT
GCTGTGTTACGGCATGTCTGTGTTCCCTCACTGGCATCGACTGTTGACGGTTCAGAGTGAAAATG
CTTTGAGACGACATGGCTACGATGGAGCTTTGCCGTACTGGGATTGGACCTCTCCTCTTAATCACC
TTCCCGAACTGGCAGATCATGAGAAGTACGTCGACCCTGAAGATGGGGTAGAGAAGCATAACCCTT
GGTTCGATGGTCATATAGATACAGTCGACAAAACAACAACAAGAAGTGTTCAGAATAAACTCTTCG
AACAGCCTGAGTTTGGTCAATTATACAAGCATTGCCAAACAAGTACTGCTAGCGTTGGAACAGGACA
ATTTCTGTGACTTTGAAATCCAATATGAGATTGCCATAACTACATCCATGCACCTGTAGGAGGCG
CTCAGCCTTATGGTATGGCATCGCTTCGCTACACTGCTTTTGATCCACTATTCTACTTGATCACT
CTAATACAGATCGTATATGGGCAATATGGCAGGCTTTACAGAAGTACAGAGGAAAACCGTACAACG
TTGCTAACTGTGCTGTTACATCGATGAGAGAACCCTTTGCAACCATTGTCCTCTCTGCCAATATCA
ACACAGACCATGTAACCAAGGAGCATTCAAGTCCATTCAACGTTTTTTGATTACAAGACCAATTTCA
ATTATGAATATGACACTTTGGAATTTAACGGTCTCTCAATCTCTCAGTTGAATAAAAAGCTCGAAG

Fig. 8b

CGATAAAGAGCCAAGACAGGTTCTTTGCAGGCTTCCTGTTATCTGGTTTTCAAGAAATCATCTCTTG
TTAAATTCAATATTTGCACCGATAGCAGCAACTGTCACCCCGCTGGAGAGTTTTACCTTCTGGGTG
ATGAAAACGAGATGCCATGGGCATACGATAGAGTCTTCAAATATGACATAACCGAAAAACTCCACG
ATCTAAAGCTGCATGCAGAAGACCATTCTACATTGACTATGAAGTATTTGACCTTAAACCAGCAA
GCCTGGGAAAAGATTTGTTCAAGCAGCCTTCAGTCATTCATGAACCAAGAATAG

INTRON 1C/1D (SEQ ID NO:140)

GTACTTGTTATATGTTTCGAATATTGCCGATACCTTCAATATATATACTTTATCAAAGTAATTGAT
TAATCTGAAGTAATTTTCCCTTCCAGTAGAGATTGAGTTGATACAACAAGAATTCGCCCTGTTGTA
TGTCACCTTTATTTTCATCAAACGATTGGAAGTGAGCTGTCCATGCCACAATGGGGTCTCTGTAAC
TTCTCGTATGGGGTATAGATTATATAGACGTGGCAGACCTTACGTATAACTAATATTTGTGTAATG
TCGTTTCAG

DOMAIN 1D

GTCACCATGAAGGCGAAGTATATCAAGCTGAAGTAACTTCTGCCAACCGTATTGAAAAAACATTG
AAAATCTGAGCCTTGGTGAACCTCGAAAGTCTGAGAGCTGCCTTCCTGGAAATTGAAAACGATGGAA
CTTACGAATCAATAGCTAAATTCATGGTAGCCCTGGTTTGTGCCAGTTAAATGGTAACCCCATCT
CTTGTTGTGTCCATGGCATGCCAACTTTCCCTCACTGGCACAGACTGTACGTGGTTGTCTGTTGAGA
ATGCCCTCCTGAAAAAAGGATCATCTGTAGCTGTTCCCTATTGGGACTGGACAAAACGAATCGAAC
ATTTACCTCACCTGATTTTCAAGACGCCACTTACTACAATTCAGGCAACATCACTATGAGACAAACC
CATTCATCATGGCAAAATCACACACGAGAATGAAATCACTACTAGGGATCCCAAGGACAGCCTCT
TCCATTCAGACTACTTTTACGAGCAGGTCCTTTACGCCTTGGAGCAGGATAACTTCTGTGATTTCTG
AGATTGAGTTGGAGATATTACACAATGCATTGCATTCTTTACTTGGTGGCAAAGGTAAATATTCCA
TGTCAAACCTTGATTACGCTGCTTTTGTATCCTGTGTTCTTCTTCATCACGCAACGACTGACAGAA
TCTGGGCAATCTGGCAAGACCTTCAGAGGTTCCGAAAACGGCCATACCGAGAAGCGAATTGCGCTA
TCCAATTGATGCACACGCCACTCCAGCCGTTTGTATAAGAGCGACAACAATGACGAGGCAACGAAAA
CGCATGCCACTCCACATGATGGTTTTGAATATCAAAACAGCTTTGGTTATGCTTACGATAATCTGG
AACTGAATCACTACTCGATTCCCTCAGCTTGATCACATGCTGCAAGAAAGAAAAAGGCATGACAGAG
TATTCGCTGGCTTCCTCCTTCACAATATTGGAACATCTGCCGATGGCCATGTATTTGTATGTCTCC
CAACTGGGGAACACACGAAGGACTGCAGTCATGAGGCTGGTATGTTCTCCATCTTAGGCGGTCAA
CGGAGATGTCCTTTGTATTTGACAGACTTTACAACTTGACATAACTAAAGCCTTGAAAAAGAACG
GTGTGCACCTGCAAGGGGATTTGATCTGGAAATTGAGATTACGGCTGTGAATGGATCTCATCTAG
ACAGTCATGTATCCACTCTCCCACTATACTGTTTGAGGCCGGAACAG

INTRON 1D/1E (SEQ ID NO:141)

GTAATATTTTGTCACTGTAACCAACAACCTGCAGTCTATTTTGCAATTACGATAATAACAATTTTT
GAAATATATCTTTATTAAAGCAAAGGTTTCTAGAGACAAACAGCCGGCTCTAATTATTTTTTCGAA
CTTACGCTTGAGTAAAGATCTGCAAATGGCAACCCTACCTATACTATTAAAAATATAATGTTACAT
TCGTATCTGAATGTTTAATAAATCACTTCATATTCTGTTGCAG

DOMAIN 1E

ATTCTGCCCACACAGATGATGGACACACTGAACCAGTGATGATTTCGCAAAGATATCACACAATTGG
ACAAGCGTCAACAACCTGTCAGTGGTGAAGCCCTCGAGTCCATGAAAGCCGACCATTTCATCTGATG
GGTTCCAGGCAATCGCTTCCTTCCATGCTCTTCCTCCTCTTTGTCCATCACCAGCTGCTTCAAAGA
GGTTTGCGTGCTGCGTCCATGGCATGGCAACGTTCCCAACAATGGCACCGTCTGTACACAGTCCAAT
TCCAAGATTCTCTCAGAAAACATGGTGCAGTCGTTGGACTTCCGTACTGGGACTGGACCCTACCTC

Fig. 8c

GTTCTGAATTACCAGAGCTCCTGACCGTCTCAACTATTTCATGACCCGGAGACAGGCAGAGATATAC
CAAATCCATTTATTGGTTCTAAAATAGAGTTTGAAGGAGAAAACGTACATACTAAAAGAGATATCA
ATAGGGATCGTCTCTTCCAGGGATCAACAAAAACACATCATAACTGGTTTATTGAGCAAGCACTGC
TTGCTCTTGAACAAACCACTACTGCGACTTCGAGGTTTCAGTTTGAAATTATGCATAATGGTGTTT
ATACCTGGGTTGGAGGCAAGGAGCCCTATGGAATTGGCCATCTGCATTATGCTTCCTATGATCCAC
TTTTCTACATCCATCACTCCCAAACCTGATCGTATTTGGGCTATATGGCAATCGTTGCAGCGTTTCA
GAGGACTTTCTGGATCTGAGGCTAACTGTGCTGTAAATCTCATGAAAACCTCCTCTGAAGCCTTTCA
GCTTTGGAGCACCATATAATCTTAATGATCACACGCATGATTTCTCAAAGCCTGAAGATACATTG
ACTACCAAAGTTTGGATACATATATGACACTCTGGAATTTGCAGGGTGGTCAATTCGTGGCATTG
ACCATATTGTCCGTAACAGGCAGGAACATTCAGGGTCTTTGCCGGATTCTTGCTTGAAGGATTTG
GCACCTCTGCCACTGTTCGATTTCCAGGTCTGTTCGCACAGCGGGAGACTGTGAAGATGCAGGGTACT
TCACCGTGTTGGGAGGTGAAAAAGAAATGCCTTGGGCCTTTGATCGGCTTTACAAGTACGACATAA
CAGAAACCTTAGACAAGATGAACCTTCGACATGACGAAATCTTCAGATTGAAGTAACCATTACAT
CCTACGATGGAACGTACTCGATAGTGGCCTTATTCCACACCGTCAATCATCTATGATCCTGCTC
ATC

INTRON 1E/1F (SEQ ID NO:142)

GTAAGTATACACACATTATTTCTCTTCTGCTATATCAGATGAAGAGAACGTTGTATCACTAACCTA
GTCTTGTTTGATTTGTGGTTTCGTTTGCTTCCTGAACAGTAGGGTTGATTTAACTTCTCTGTTTCG
TCTGTACCAATGAAAGACTATGATGCTTGTGTGAAGATGCTTTGTTTCATGAGTCAGTCTGTTCTTG
TAATGCTTTGATCTTTGCCATCAACATTCTTGAAATTAATTATGGTTTCCCTTAAATACTTACATA
TTACATTTAAACGTCGCTGCTTGTCTGATTGCATATTCTTTCAAAAATAACTATATATTCAG

DOMAIN 1F-1 (1st part of domain f)

ATGATATTAGTTCGCACCACCTGTCGCTCAACAAGGTTTCGTTCATGATCTGAGTACACTGAGTGAGC
GAGATATTGGAAGCCTTAAATATGCTTTGAGCAGCTTCGAGGCAGATACCTCAGCAGATGGTTTTG
CTGCCATTGCATCCTTCCATGGTCTGCCTGCCAAATGTAATGACAGCCACAATAACGAG

INTRON 1F-1/1F-2 (SEQ ID NO:143)

GTAAATATACAGTGAAATCCGGATAAGTAAATCCAGATAAGAAAAAAACATTTTCTGTGGTCCC
GGCATGTTTCTTCTTCATCTATCATTATTTTGATACGGATAAGTAAAAATCGGCTGAGTAAACAT
CCGGGTAAGTAAATGATTTTCGAGGTCTCTTCATCGGATAAGTAAAGATACACAAGTGATCATTCC
AATAAACACTAACTGATGCAACACAATACCAGCGCACAGTGTTTTCACTACGTTTGTGTATTGT
AATTAACAATTAACACTTAAGTGTTTCCCAATGTGTCCGTGTGCAAACTGATTGGGACAAAGCTTG
CAACAAGCCCGCAATTCCATGTCGTTTATGTCTACGTTTGTATTCTGACTGCTTGGAGGGGTTT
GGAAAAAATAAAAAACGGGTAAATATTATAAAAAATTCACGGTGCCTTGAAATTTTAGGTGTCCG
GATTTCACTGTAGATGATTAATTTCTCACTTGTAACAAAAGGACCCAGTACCCTCATTCGTGAC
GTACGTTATAAAATGTAATTATAAAAAGCCATTATCATGTTATACGTGATCTTGNCTTGCAATTA
TNCTACCGCTTTCTTGATTTTTTAAAGCAATTTCTCCCTCTATGAACTTATTAACATAGCACTCCT
GCAAAAGAAAACAGTCACTGCATGGATCCATATTGAATGTTGCTGCTTATTTCTCATTTTATTACT
CACAGATATTTCAAGAACATCGTACTCTTAACCAGGCTAAAGCAAAGAGGGTTACATTTTAGCCG
ACAAGTTCAGTAGCTGAGTGGAACACGTATATTAATGGAGATGACTCTGGTCATGATGATTAGG
ACAATTATCATGACGTTATCATTGATCATGACCATGTCAGTATAATAGATAGCTAACAAATAATGT
AATTACTAATTATGAAGCAATGGTGCATTTGCAG

DOMAIN 1F-2 (2nd part of domain f)

Fig. 8d

GTGGCATGCTGTATCCATGGAATGCCTACATTCCCCCACTGGCACAGACTCTACACCCTCCAATTT
GAGCAAGCTCTAAGAAGACATGGCTCTAGTGTAGCAGTACCCTACTGGGACTGGACAAAGCCAATA
CATAATATTCCACATCTGTTTCACAGACAAAGAATACTACGATGTCTGGAGAAATAAAGTAATGCCA
AATCCATTTGCCCCGAGGGTATGTCCCCTCACACGATACATACACGGTAAGAGACGTCCAAGAAGGC
CTGTTCCACCTGACATCAACGGGTGAACACTCAGCGCTTCTGAATCAAGCTCTTTTGGCGCTGGAA
CAGCAGCACTACTGCGATTTTGCAGTCCAGTTTGAAGTCATGCACAACACAATCCATTACCTAGTG
GGAGGACCTCAAGTCTATTCTTTGTCATCCCTTCATTATGCTTCATATGATCCGATCTTCTTCATA
CACCCTCCTTTGTAGACAAGGTTTGGGCTGTCTGGCAGGCTCTTCAAGAAAAGAGAGGCCTTCCA
TCAGACCGTGCTGACTGCGCTGTTAGTCTGATGACTCAGAACATGAGGCCTTTCCATTACGAAATT
AACCATAACCAGTTTCACCAAGAAACATGCAGTTCCAAATGATGTTTTCAAGTACGAACTCCTGGGT
TACAGATACGACAATCTGGAAATCGGTGGCATGAATTTGCATGAAATTGAAAAGGAAATCAAAGAC
AAACAGCACCATGTGAGAGTGTTTGCAGGGTTCTCTCCTTCACGGAATTAGAACCTCAGCTGATGTC
CAATTCCAGATTTGTAAACATCAGAAGATTGTCACCATGGAGGCCAAATCTTCGTTCTTGGGGGG
ACTAAAGAGATGGCCTGGGCTTATAACCGTTTATTCAAGTACGATATTACCCATGCTCTTCATGAC
GCACACATCACTCCAGAAGACGTATTCCATCCCTCTGAACCATTCTTCATCAAGGTGTCAGTGACA
GCCGTCAACGGAACAGTTCTTCCGGCTTCAATCCTGCATGCACCAACCATTATCTATGAACCTGGT
CTCGGTG

INTRON 1F-2/1G-1 (SEQ ID NO:144)

GTCTCGGTGAGTTATTAAAAGAAACAAAATATTTACCATTACCATTGTAACTACAAAAATGAGTG
AGATATCTTATATCACTGGTACACTACTGATATTTTATGCAATGAAATTACTATTTTTTCCAGGTAC
GCTTCAACCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCATCATGCTTTTCTGT
AAAACATAAAACACCAATTAACAATGTTCTTAGTGTGTTTGTGACTCCCTTCCACTGCAACGCCT
ACATAATCAAAGTGTTCTGTTTTTTTCCAACTTTCCAGTTAGTGTGTTGAAGACTAAAAAGTTAAATA
AGCATTACATAACTTCTAAGAGCAACTGGGACCATGCAGTTACGTATTGATATTTCTGTGAGAGT
GAAGCAAAACACTGTTTTTTCAAGCTTAGGTTTATCAATCAAATGTCCAATAGTTCATGTTATCGA
AAAGGCAGCGAAGGATAAGAGGCTCCGAGACATCTTGTCTATTCTCGTGTTCATATGATATCAACT
GAGGAGCTTCCATTACATTTTTTGACCTTATCATTTAAAGACATACATGGAACATTTTCATTTTACA
GTTAAAGTGAACCACTTCAGGTTCAACTTCAACTTCGAATTCAACTTCTGTTGTGTGTTTTATGAG
CCGACTGAAATAGAGTGCCTTACTTTCACTTCTAGTTTCGTTCTGTCTCGTCATCGTTGTTTCTTT
CAGTGTGCATAGTACACGCCTAGTATAGAACACACGAACCTTGTCTTACTTAATAGATTCTGAAAC
TATTATGTGGAAAGTTGGCAGGCTATAGTAACATCCTGGCAAAATTATCATGTATCCTCTTGTTTG
TCATAATTAG

DOMAIN 1G-1 (1st part of domain g)

ACCATCACGAAGATCATCATTTCTTCTTCTATGGCTGGACATGGTGTGAGAAAGGAAATCAACACAC
TTACCACTGCAGAGGTGGACAATCTCAAAGATGCCATGAGAGCCGTGATGGCAGACCACGGTCCAA
ATGGATACCAGGCTATAGCAGCGTTCCATGGAAACCCACCAATGTGCCCTATGCCAGATGGAAAGA
ATTACTCGTGTGTGACACATG

INTRON G1-1/1G-2 (SEQ ID NO:145)

GTATGTATTTCCCACTGGTGGTCTGCTGACTGCCAACACATACTTGTAATTTATTCATGAAAGTATA
ATAGTTTGTGTTGAAAGTATATTTATAACCATCTTGACACAAGCGTCACGAATTTTACCACAAAGCT
TCAAAACGCCCAAAACATTCTAATAGCGATATATTTGTTAAAAGACCAAAATATAGCCTTACAACA
ATAGATTATTTTAATAAGACCAGTCAGTGCATGCAATCGATTGGAAACTTTGAAATAAAATATTC
TATGTACTAACTGCCAATCTCATAATACTTGCCTTGATGTGCTTCTTTTTTCACATTCGCGTCGAG
CTTCAACTCCAATGCATAAGCTTAAAAATAATCATAAACACAAACAAATAGCCACAGAGGCGACGA
TCCCTCCAGGCCAGGCTTTATTTGTCTCTTATAGAATATATCGCTATTAGAATGTTTTTGACGTTT

Fig. 8e

TGAAGCTTTGTGGGTGAAAATTCGTGATGTTTATGCGTGGTATTTATGTAAGATGAAAATAAATAT
ATCTTTTCAAACAAGATTTTAGTATTTTGAAGACTTCTATGAATAAATTACACTTATGTGTTAGGT
TATTGGTCACTGAGCGCTTGTGGTATTTTCCCTTCTTCAATTTGTTTGTTCCTTTGTTCAATTTCGA
ATAGTTATCCTACTGTGGATAGTCTATATGAGAATCGTTGAAAGAATAATACAATTCTAATGGATT
GCAACTTCTTTAACTTTTATTTGCAACTGCCACGTTTCGGTATACGTTCTTATGCCGTCATCAAGC
ATACGAGTGTACATGTATGCCAAAACGCTGCAATAAAAATTAAAGAAGTTGCAATCCATAAGAAT
TTCAATGTTCTTTTCATCATCACATCAACTTCTAAAAATGCCTATAAAACAATCAACAAACGTACAA
TAGTACATTACCGGATCTCGCAGCATGACCACGTCGATATCTAAACAATATCACTATCCATTAATA
GGATCAAGAGTAGGTACAGACATGTTTCAAGTTATAAATACTCTTCAAAAAAGTAGGGGAACTTGGA
TTTCAAGGTCAATAACAACTAATGATAATAACAATTGGTCCCAAATAATAACAATTGGTCCCAA
CTAATTGTATCTTTACAAAGAAGAAATTGAGTGAACAATTCACCCGGTATTTTATTACCTAAACCG
TTTCTCTTGCTGTTATGGTGCGTGAAAGAAGAAATGGGTAAGAAACGGAAATTGACATTTTTCGCT
CAGTGGTGCGTAATGCCCCCATTGTTGGCCAAACACTGATTGATTGCTGAGGCATCGTGCATACG
CGTCTACCTATGGTAATTTGATGCAGTCTGTCCATTCTTCCACCAACGCCTGGACAAGTTCATCT
AGCGTGGCTGGTGGCCTTTACGTTGACGCACACGTCGGCCCAAGATGTCCAGACATTTTCAATG
GCCAGGGCTCATTGCTGGTCAGGGCATCCTATGGATATTGTGCCGTTGAAGGTGGTTATGTTGTTT
ACATTGAAATTCCAAGTTCTCCTACTCTTTTAAAGAGGAGGTTACAAAGTACGTTCTTTTCATGTT
GGTGAAGAGAATATCAAGGTCTTCTAAGGGATTGTGTCTTATAATATTTGATTTTAAAGAAGTTTGA
TATTATCTGCATCCTTCCCAAGAAATTGCAATGTTTACACACTATTGCGTTTGATAATGTTTTTG
GGGAAATAAAGTGTCCAGGACTGCTAAATAGTAATTATTGCTACTTTTAG

DOMAIN 1G-2 (2nd part of domain g)

GCATGGCTACTTTCCCCCACTGGCACAGACTGTACACAAAACAGATGGAAGATGCCTTGACCGCCC
ATGGTGCCAGAGTCGGCCTTCTTACTGGGACGGGACAACCTGCCTTTACAGCTTTGCCAACTTTTG
TCACAGATGAAGAGGACAATCCTTTCCATCAT

INTRON 1G-2/1G-3 (SEQ ID NO:146)

GTGAGTTCACGTAAGCCTACGAGATCAACATTACTCCTTAACAGCCACGGCATCATGTACCGATAT
ATCACAACAAAAGTATTCAAAGCTTTAAACACGATATGTATGGTTCAGAATGACATCATTAAC
AAGGACATGAGTCTGAAATAAACATGACTTGACACCGTTGTGGTCACAGTTTTGTTTCTCATTGGT
GAACCTGTGAAACAACCTTTCAAACCAAAGATGCCTATTAATATTGTTAATTCCCATGAATTAGG
AGATACACACATTCTACTGTCATTT.....AATAACCGCTTC
CAGCATGAAAACACAATATGATTATCTCAATTCTACCATTACTAATTATAATTTTGACTGGCATT
TTTGACGACGCGTAAACATCGCTGCTTTACAGACTGCACTGCGGTAACGTGACGTTTTTCATGAC
GTCACTACATTCTATTCAAACATTTCCACAGAAGAGCGAGACCACGGCCGTGATGGGTTCTGGGC
AGATGATTACCCAAGTATATATTTATAATAACTTGACTGCTTGCCCTGAATAATGTTGACACATGAC
AACGAATTTGTGATAGCGTAAGAAGCGTGAATACTGTGAATAGTGTGAGGGGTGTTTGCTGAGAGT
TAACCACCGTTAATTGCAAAAATTCCTGAATACCTTGCAATTTGCAGTCAAGAAGAATTGCATTCTTA
CTCCTGTGAATGGACTCATTGTTATTTAGCAGCGGTTATTGAGGTTTTGATCACCTCTAAATAGAC
AATCAGGATGCGGCAAACCGGAAAATTATAGCAGAATCTGTAATTCAAGATGGGCTTGCCCTGTGAA
AATATGCTGCGAGTTCAGTAACACTTTTCCCTTTTCGATCATGGCCTGTTTTGCTCTGAATCTGGTC
TTTCAGAGGATCCCTGCTTTTTTAAACTAAAGTCTCCCAACTCACTTATATTTATGTTTTTTAA
TTATTTATAGTTTTTAATATGAACAACAAATCATATTTATTTACACATTATATTTTTTCAG

DOMAIN 1G-3 (3rd part of domain g)

GGTCACATAGACTATTTGGGAGTGGATACAACTCGGTGCCCCGAGACAAGTTGTTCAATGATCCA
GAGCGAGGATCAGAATCGTTCTTCTACAGGCAGGTTCTCTTGGCTTTGGAGCAGACAGAT

Fig. 9a**Primary structure of the KLH1 protein****DOMAIN B**

GLPYWDWTEPMTHIPGLAGNKTYVDSHGASHTNPFHSSVIAFEENAPHTKRQIDQRLFKPATFGHH
TDLFNQILYAFEQEDYCDFEVQFEITHNTIHAWTGGSEHFSMSSLHYTAFDPLFYFHHSNVDRLWA
VWQALQMRRHKPYRAHCAISLEHMHLKPPAFSSPLNNNEKTHANAMPNKIYDYENVLHYTYEDLTF
GGISLENI EKMIHENQQEDRIYAGFLLAGIRTSANVDIFIKTDSVQHKAGTFAVLGGSKEMKWGF
DRVFKFDITHVLKDLDLTADGDFEVTVDITEVDGTKLASSLI PHASVIREHARGKLN R

DOMAIN C

VKFDKVP RSRLIRKNVDRLSPEEMNELRKALALLKEDKSAGGFQQLGAFHGEPKWCPSP EASKKFA
CCVHGMSVFPWHRLTLTVQSENALRRHGYDGALPYWDWTSPLNHLPELADHEKYVDPEDGVEKHNP
WFDGHIDTVDKTTTRSVQNKLF EQPEFGHYTSIAKQVLLALEQDNFCDFEIQYEIAHNYIHALVGG
AQPYGMASLRYTAFDPLFY LHSNTDRIWAIWQALQKYRGKPYNVANCAVTSMREPLQPFGLSANI
NTDHVTKEHSVPFNVDYKTNFN YEYDTLEFNGLSISQLNKKLEAIKSQDRFFAGFLLSGFKKSSL
VKFNICTDSSNCHPAGEFYLLGDENEMPWAYDRVFKYDITEKLHDLKLHAEDHFYIDYEVFDLKPA
SLGKDLFKQPSVIEPRI

DOMAIN D

GHHEGEVYQAEVTSANRIRKNIENLSLGELESRAAFLEIENDGTYESIAKFHGSPGLCQLNGNPI
SCCVHGMPTFPWHRLYV VVENALLKKGSSVAVPYWDWTKRIEHLPHLISDATYYNSRQH HYE TN
PFHHGKI THENEITTRDPKDSL FHSDFYEQVLYALEQDNFCDFEIQLEILHNALHSL LGGKGKYS
MSNLDYAAFDPVFFLHHATTDRIWAIWQDLQRFRKRPYREANCAIQLMHTPLQPFDKSDNNDEATK
THATPHDGF EYQNSFGYAYDNLELNHYSIQLDHMLQERKRHDRV FAGFLLHNIGTSADGHV FVCL
PTGEHTKDCSHEAGMFSILGGQTEMSFVFDRLYKLDITKALKKNGVHLQGD FDL EIEITAVNGSHL
DSHVIHSP TILFEAG

DOMAIN E

TDSAHTDDGHTEPVMIRKDITQLDKRQQLSLVKALESMKADHSSDGFQAIASFHALPPLCPSPAAS
KRFACCVHGMATFPQWHRLYTVQFQDSL RKHGAVVGLPYWDWTLPRSELPELLTVSTIHD PETGRD
IPNPFIGSKIEFEGENVHTKRDINRDLRFQGSTKTHHNWFIEQALLALEQNTNYCDFEVQFEIMHNG
VHTWVGKPEPYGIGHLHYASYDPLFYIHHSQTDRIWAIWQSLQRFRGLSGSEANCAVNLMKTPLKP
FSFGAPYNLNDHTHDFSKPEDTFDYQKFGYIYDTLEFAGWSIRGIDHIVRNRQEHSRVFAGFLLEG
FGTSATVDFQVCRTAGDCEDAGYFTVLGGEKEMPWAFDRLYKYDITETLDKMNLRHDEIFQIEVTI
TSYDGTVLDSGLIPTPSIIYDPAH

DOMAIN F

HDISSHLSL NKVRHDLSTLSE RDIGSLKYALSSLQADTSADGFAAIASFHGLPAKCND SHNNEVA
CCIHGMP TFPWHRLYTLQFEQALRRHGSSVAVPYWDWTKPIHNI PHLFTDKEYYDVWRNKVMPNP
FARGYVPSHDTYTVRDVQEGFLH LTSTGEHSALLNQALLALEQHDYCDFAVQFEVMHNTIHYLVGG
PQVYSSL SHYASYDPIFFI HHSFVDKVVAVWQALQEKRLPSDRADCAVSLMTQNMRPFHYEINH
NQFTKKHAVPNDVFKYELLGYRYDNLEIGGMNLHEIEKEIKDKQHHVRVFAGFLLHGIRTSADVQF
QICKTSEDCHHGGQIFVLGGTKEMAWAYNRLFKYDITHALHDAHITPEDVFHPSEPF FIKVSVTAV
NGTVLPASILHAPTIIYEPGLG

Fig. 9b

DOMAIN G

DHHEDHHSSSMAGHGVRKEINTLTAEVDNLKDAMRAVMADHGPNYQAIAAFHGNPPMCPMPDGK
NYSCCTHGMATFPHWHRLYTKQMEDALTAHGARVGLPYWDGTTAFTALPTFVTDEEDNPFHHGHID
YLGVDTTTRSPRDKLFNDPERGSESFFYRQVLLALEQTD

Fig. 10a**Genomic sequence of the KLH2 gene****DOMAIN 2B**

GGCCTGCCCTACTGGGATTGGACCATGCCAATGAGTCATTTGCCAGAACTGGCTACAAGTGAGACC
TACCTCGATCCAGTTACTGGGGAACTAAAAACAACCCCTTTCCATCACGCCCAAGTGGCGTTTGAA
AATGGTGTAAACAAGCAGGAATCCTGATGCCAACTTTTTATGAAACCAACTTACGGAGACCACACT
TACCTCTTCGACAGCATGATCTACGCATTTGAGCAGGAAGACTTCTGCGACTTTGAAGTCCAATAT
GAGCTCACGCATAATGCAATACATGCATGGGTTGGAGGCAGTGAAAAGTATTCAATGTCTTCTCTT
CACTACACTGCTTTTGATCCTATATTTTACCTCCATCACTCAAATGTTGATCGTCTCTGGGCCATT
TGGCAAGCTCTTCAAATCAGGAGAGGCAAGTCTTACAAGGCCCACTGCGCCTCGTCTCAAGAAAGA
GAACCATTAAGCCTTTTGCATTTCAGTTCCCCACTGAACAACAACGAGAAAACGTACCACAACCTCT
GTCCCCACTAACGTTTATGACTATGTGGGAGTTTTCACACTATCGATATGATGACCTTCAGTTTGGC
GGTATGACCATGTCAGAACTTGAGGAATATATTCAACAAGCAGACACAACATGATAGAACCTTTGCA
GGATTCTTCTTTTCATATATTGGAACATCAGCAAGCGTAGATATCTTCATCAATCGAGAAGGTCAT
GATAAATACAAAGTGGGAAGTTTTGTAGTACTTGGTGGATCCAAAGAAATGAAATGGGGCTTTGAT
AGAATGTACAAGTATGAGATCACTGAGGCTCTGAAGACGCTGAATGTTGCAGTGGATGATGGGTTT
AGCATTACTGTTGAGATCACCGATGTTGATGGATCTCCCCCATCTGCAGATCTCATTCCACCTCCT
GCTATAATCTTTGACGTGGTCAGAG

INTRON 2B/2C (SEQ ID NO:147)

GTATTTAAAAAAGTAATAAAACCATATTTTTCGAATGCGCTTTATGAAATATCGTGTGACTGGTTCT
TTAGTTTACATGGAGTGTAAACAACATGCTCCATCAGTTGACATATACTGCTCACACAAAGTAAGGG
ATATTTGATAATGATAACAAATATAATCAAAGCGGTTATACTATCAAGACTTATTCACATAATTAC
AGGTGAAGGGAGGTGTGATCGTGTTCACTGATCAGGTTGAGGCCAGAGAAGTCCCAGTTTGAGTCT
TGCAGAAGATGATGTTTAGGCATGGGGTTCGAATCACCAAAATCACATGACTTCAATAACGGGTGG
ACCACCTCGAGCGACGATGCAAGCAGTAGAGCGTCTACGCATGCTCCTGATAAGGCGACCAATCTG
TTCTGTTGGGAATCAGTCGCCACTCCTCTTGTAGTGCCACGCTCATTTCTGCTACGGTCTGGGTAC
CTGCTATCGGGTCTTGATCCGTATCCCAAGGATGTCCACACATGTTCAAGGTGAGAGGTTCGGGGA
ACATCGCTGGCCACGGTAAGGTCTGAATTTGATGCCGTTGAAAGTGAGCTCTGACAACCTGAGCAT
GGTGAGCTCTGACGTTGTCTGCTGAAAGATGAATCCAGCTCCATGACAGCGAGCAAAGGGCAGGA
CGTGTTGGTCAATGCAGTTGTCTCTGCAGTACACACCTGTCACTCGCCACTCACAAGCGTGTAGAT
CTGTACGACCAGTCATGGAGATCCCAGCCCACATCATAACGGACCCCTATCCATACCGATCATGAG
CCACCATAGCAGCGTCTTGATGACGTTCTCCCTGTGCGCTCGACATCCTCACACGGCCAAAAGGAA
CGTGGACTCGTCACTGAACATGACATTAGCCAACCTGGCACTTGTCCACCGCTGATGTTGGCGAGA
CCATTCCAGTCGAGCTCTTCGGTGTCTGGCTTTCATCGATAAACACGACGTAAGGTCTGCGGGCGTG
CAAGACGGCTCTATGCAGGCGATTTTCGGATTGTCTGGGTGCTAACTCTGATCCCAGGTGCCTGCTG
AAGTTGATGCTGGATCTGTGTGGCATTGAGATGGCGATTCCCTTAGGACTGTGGAGATGATGAATCG
ATCTTGACTTATGGTGGTGACATTAGGACGTCGGGTTCTGTCTCTATCCTGCACTCTTCCAGTTGT
TCGGTGACGCTCTGGTACCCGGCTGATTACTGACTGAGAATATCCATCTGCCGTGCGACATGAGCC
TGTGTTGGCCCAGCCTGAAGCATTGCAATCGCCAGAGACGCTCTTCAAAGTCAATTCGACGCATGG
TTTTCTGTTTCAAAATGACAGCGTAAACAGTTTTTGGTGCTTTTATGCTTCCCAAGAGCATGAAA
AACACGTTCTATGGGTCGTGCACACCTTACATGACAAGTGTGAAAAGTGACTTGCACCCCCCTTGTG
TGTTCCGGATGCACACTCTGTTTACGTACTGATGCGATTTGGCGTCTAAACATGTTTTGGCGTCTAA
ACATGTTTTCTGTCATGATTCATATACTATTTTGTTCATATTCCTGGCATCAAACCAAACCTACAGTG
AAATATATTTCAATATCCCCTACTTTGTGTGAGTAGTATAGATCACTGCAGACAACATATAGACAA
TGCAGTTACACCGTCAACAATCCCAGTCATTAATTATGATGACACTTCCACACATAGTGTCAAGTGA
TTGTAATTCAACTGTACACACTTTTCCCGTGAACATTCAGGATCTATATGACTAAATATATAACAT
TAGTATACGTGCAGTTTTGTATCGCTACGACATTGTTGTAACCTTTGTTTAAATCATTTAACAG

Fig. 10b

DOMAIN 2C

CTGATGCCAAAGACTTTGGCCATAGCAGAAAAATCAGGAAAGCCGTTGATTCTCTGACAGTCGAAG
AACAAACTTCGTTGAGGCGAGCTATGGCAGATCTACAGGACGACAAAACATCAGGGGGTTTCCAGC
AGATTGCAGCATTCCACGGAGAACC AAAATGGTGTCCAAGCCCCGAAGCGGAGAAAAAATTTGCAT
GCTGTGTTTCATGGAATGGCTGTTTTCCCTCACTGGCACAGATTGCTGACAGTTCAAGGAGAAAATG
CTCTGAGGAAACATGGATTTACTGGTGGATTGCCCTATTGGGACTGGACTCGGCCAATGAGCGCCC
TTCCACATTTTGTGCTGATCCTACTTACAATGATTCTGTTTCCAGCCTCGAAGAAGATAACCCAT
GGTATCATGGTTCACATAGATTCTGTTGGGCATGATACTACAAGAGCTGTGCGTGATGATCTTTATC
AATCTCCTGGTTTTCGGTCACTACACAGATATTGCAAAACAAGTCCTTCTGGCCTTTGAGCAGGACG
ATTTCTGTGATTTTGGAGGTACAATTTGAAATTGCCATAATTTTCATACATGCTCTGGTTGGTGGTA
ACGAACCATAACAGTATGTCATCTTTGAGGTATACTACATACGATCCAATCTTCTTCTTGACCGCT
CCAATACAGACCGACTTTGGGCCATTTGGCAAGCTTTGCAAAAATACCGGGGGAAACCATAACA
CTGCAAACTGTGCCATTGCATCCATGAGAAAACCACTTCAGCCATTTGGTCTTGATAGTGTCTATA
ATCCAGATGACGAACTCGTGAACATTCGGTTCCTTTCCGAGTCTTCGACTACAAGAACAACCTTCG
ACTATGAGTATGAGAGCCTGGCATTTAATGGTCTGTCTATTGCCCAACTGGACCGAGAGTTGCAGA
GAAGAAAGTCACATGACAGAGTCTTTGCAGGATTTCCTTCTTCATGAAATTGGACAGTCTGCACTCG
TGAAATTCTACGTTTGCAAACACAATGTATCTGACTGTGACCATTATGCTGGAGAATTCTACATTT
TGGGAGATGAAGCTGAGATGCCTTGGAGGTATGACCGTGTGTACAAGTACGAGATAACACAGCAGC
TGCACGATTTAGATCTACATGTTGGAGATAATTTCTTCCTTAAATATGAAGCCTTTGATCTGAATG
GCGGAAGTCTTGGTGGAAAGTATCTTTTCTCAGCCTTCGGTGATTTTCGAGCCAGCTGCAG

INTRON 2C/2D (SEQ ID NO:148)

GTATGTTTTAAATGTCACCTTATCCGTGATCTGTAATGAAGTTAGCAATTCACCTTTATCAACTGTTT
GGCTGTACTGTTTCAGTGCGAGTTTTACTTAGGTTGGATTAATTAAAATATTCAAGCTCATAAATG
TTTTGATTCAACTTTTGTATTTATTTCAAACAG

DOMAIN 2D

GTTACACCAGGCTGATGAATATCGTGAGGCAGTAACAAGCGCTAGCCACATAAGAAAAAATATCC
GGGACCTCTCAGAGGGAGAAATTGAGAGCATCAGATCTGCTTTCCTCCAAATTCAAAAAGAGGGTA
TATATGAAAACATTGCAAAGTTCCATGGAAAACCAAGGACTTTGTGAACATGATGGACATCCTGTTG
CTTGTTGTGTCCATGGCATGCCCACCTTTCCCCACTGGCACAGACTGTACGTTCTTCAGGTGGAGA
ATGCGCTCTTAGAACGAGGGTCTGCAGTTGCTGTTCTTACTGGGACTGGACCGAGAAAGCTGACT
CTCTGCCATCATTAATCAATGATGCAACTTATTTCAATTCACGATCCCAGACCTTTGATCCTAATC
CTTCTTCAGGGGACATATTGCCTTCGAGAATGCTGTGACGTCCAGAGATCCTCAGCCAGAACTAT
GGGACAATAAGGACTTCTACGAGAATGTCATGCTGGCTCTTGAGCAAGACAACTTCTGTGACTTTG
AGATTACGCTTGAGCTGATACACAACGCCCTTCATTCTAGACTTGAGGGAAGGGCTAAATACTCCC
TTTCGTCTCTTGATTATACCGCATTTGATCCTGTATTTTTCTTCCATTGCAACGTTGACAGAA
TCTGGGCCATCTGGCAGGACTTGCAGAGATATAGAAAGAAACCATACAATGAGGCTGACTGCGCAG
TCAACGAGATGCGTAAACCTCTTCAACCATTTAATAACCCAGAACTTAACAGTGATTCCATGACGC
TTAAACACAACCTCCCACAAGACAGTTTTTGATTATCAAAACCGCTTCAGGTACCAATATGATAACC
TTCAATTTAACCCTTCAGCATACAAAGCTAGACCAAATTTAGGCTAGAAAACAACACGACA
GAGTTTTTGTGCTGGCTTTATTTCTTCAACATTGGGACATCTGCTGTTGTAGATATTTATATTTGCG
TTGAACAAGGAGGAGAAACAAAACCTGCAAGACAAAGGCGGGTTCCTTCACGATTCTGGGGGGAGAAA
CAGAAATGCCATTCCACTTTGACCGCTTGTAACAATTTGACATAACGTCTGCTCTGCATAAACTTG
GTGTTCCCTTGGACGGACATGGATTTCGACATCAAAGTTGACGTGAGAGCTGTCAATGGATCGCATC
TTGATCAACACATCCTCAACGAACCGAGTCTGCTTTTTGTTTCTGGTGAACGTAAGAATATATATT
ATG

Fig. 10c

INTRON 2D/2E (SEQ ID NO:149)

GTTATAAAGCAGTATATTCTCTTCAAAAAAGTAGGGGAACCTTGGAAATTTCAAGGTAAATAACATAA
CTACCTTCAACGGCACAATATCCATATGATGCCCTGGCCAGCAATGAGGCCTGATCTTTTCCCAT
TAAAAATGTCTGGAACATCTTGGGCAAACGTGTGCGTCAACGTAAACGCCACCAGTCACGCTAGA
TGAACCTTGTCCAGGCGTTGGTGGGAAGAAATGGGACAGACTGCATCAATTACCATAAGTAGACTCATT
TGCAGCGAATCAGTCAGTGTGTTGACCAATAACGGGGGCATTACGCACTACTGACGCAAAACAATGT
CAATTTCCGTTTCTTACCCATTCCCTTCTTTCACGGACCATAACAGCAAGAGAACTGNTTAGGTAA
TGAAATACCGGTGAATTATTGTTAACTGGATTCCCTTCTTTGTAAAGATAACAATTAGTTTGGGACCA
ATTATTATTATCATTAGTTTGTATTGACCTTGAAATTCGAAGTTCCTCTACATTTTTTAAGGAGT
TTATTTGATTGACAATGAAATGTAAGAAAAGAGCAAATCGTAAATACGTTAAAAATTATTCCTTA
AACATCAGTCTCTAACTTCAGTTTAAATTGCCAGTAACACGTGTTATATGATGTTTCCGTTTCTCT
TTGTTTTTTTAGCATTCAACTTATTTGATATAACGTTTTACTGTTTTAGATTACATCAAACCTGCAG

DOMAIN 2E

ATGGGCTTTTCAACAACATAATCTTGTGCGAAAAGAAGTAAGCTCTCTTACAACACTGGAGAAACATT
TTTTGAGGAAAGCTCTCAAGAACATGCAAGCAGATGATTCTCCAGACGGATATCAAGCTATTGCTT
CTTTCCACGCTTTGCCTCCTCTTTGTCCAAGTCCATCTGCTGCACATAGACACGCTTGTTGCCTCC
ATGGTATGGCTACCTTCCCTCAGTGGCACAGACTCTACACAGTTCAGTTCGAAGATTCTTTGAAAC
GACATGGTTCTATTGTCTGGACTTCCATATTGGGATTGGCTGAAACCGCAGTCTGCACTCCCTGATT
TGGTGACACAGGAGACATACGAGCACCTGTTTTACACAAAACCTTCCCAAATCCGTTCTCTCAAGG
CAAATATAGAATTTGAGGGAGAGGGAGTAACAACAGAGAGGGATGTTGATGCTGAACACCTCTTTG
CAAAGGAAATCTGGTTTACAACAACCTGGTTTTGCAATCAGGCACTATATGCACTAGAACAAGAAA
ATTACTGTGACTTTGAAATACAGTTCGAAATTTTGCAATAATGGAATTCATTCATGGGTTGGAGGAT
CAAAGACCCATTCAATAGGTCATCTTCATTACGCATCATAACGATCCACTGTTCTATATCCACCATT
CGCAGACAGATCGCATTGTTGGGCTATCTGGCAAGCTCTCCAGGAGCACAGAGGTCTTTTCAAGGAAAGG
AAGCACACTGCGCCCTGGAGCAAATGAAAGACCTCTCAAACCTTTCAGCTTTGGAAGTCCCTATA
ATTTGAACAAACGCACTCAAGAGTTCTCCAAGCCTGAAGACACATTTGATTATCACCGATTCCGGT
ATGAGTATGATTCCCTCGAATTTGTTGGCATGTCTGTTTCAAGTTTACATAACTATATAAAACAAC
AACAGGAAGCTGATAGAGTCTTCGCAGGATTCTTCTTAAAGGATTGGAACAATCAGCATCCGTAT
CGTTTGATATCTGCAGACCAGACCAGAGTTGCCAAGAAGCTGGATACTTCTCAGTTCTCGGTGGAA
GTTTCAAGAAATGCCGTGGCAGTTTGGACAGGCTTTACAAGTACGACATTACAAAAACGTTGAAAGACA
TGAAACTGCGATACGATGACACATTTACCATCAAGGTTACATAAAGGATATAGCTGGAGCTGAGT
TGGACAGCGATCTGATTCCAACCTCTTCTGTTCTCCTTGAAGAAGGAAAGC

INTRON 2E/2F (SEQ ID NO:150)

GTATGTATCTCATGTTTCTCAAATAATTTGATTTTCAATGCCCTTACTATAAAGCACAGTTATTGT
TCAGTGCCAGTAACCGTTTATTTACGTAAATGTTACAGGCTATTATAATCAAAAATACATTACCGA
TATTGTTTACCACACAATTATATCATTGTCAAAATCTACCCCCATTACCTGCGTTTTGAATTTGTA
ACCTTCTGACAAAAATGAATTAGCAAGAGCTCTGATGAAGAACATAATGAACAACACCTATCTTTC
TTCTTTCAATGACGGTTTAACAATACAATGCACAATGTAAAAAATATATATATATATAATTTT
ATATCTACAGTTAATGCAAATGACTCCACTAATTCAGGGAAACACATTTTCAG

DOMAIN 2F-1 (1st part of domain f)

ATGGGATCAATGTACGTCACGTTGGTCGTAATCGGATTCGTATGGAACCTATCTGAACTCACCGAGA
GAGATCTCGCCAGCCTGAAATCTGCAATGAGGTCTCTACAAGCTGACGATGGGGTGAACGGTTATC
AAGCCATTGCATCATTCCACGGTCTCCCGGCTTCTTGTTCATGATGATGAGGGACATGAG

Fig. 10d

INTRON 2F (SEQ ID NO:151)

GTAAATAAAACGTCCAGTCATCGGAAACCCGCCAGATATATGGGTTTTTTTCTATTTAAACAAA
AAAGCAGAGACAAAAGATTATTTAAAGTCACATTTAACTTGATATCAGATCAATAGTTTGGCTAG
TTAGTGCTCTATATCCCTCAAATCCTTCGAATCTTTAAGCCTCGTGATATTTTGACAAACAGAGAA
GACTTAGTAGCCCAGACTTTCCTTATTTTTTCCTGAAAATCTTAATACGGATATTAAATGGATTTC
ATTCTGCAACCTACAACCATAGCCCATATGTTATTATTTTCAG

DOMAIN 2F-2 (2nd part of domain f)

ATTGCCTGTTGTATCCACGGAATGCCAGTATTCCCACACTGGCACAGGCTTTACACCCTGCAAATG
GACATGGCTCTGTTATCTCACGGATCTGCTGTTGCTATTCCATACTGGGACTGGACCAAACCTATC
AGCAAACCTGCCTGATCTCTTCACCAGCCCTGAATATTACGATCCTTGGAGGGATGCAGTTGTCAAT
AATCCATTTGCTAAAGGCTACATTAAATCCGAGGACGCTTACACGGTTAGGGATCCTCAGGACATT
TTGTACCACTTGCAGGACGAAACGGGAACATCTGTTTTGTTAGATCAAACCTCTTTTAGCCTTAGAG
CAGACAGATTTCTGTGATTTTGAGGTTCAATTTGAGGTCGTCCATAATGCTATTCACTACTTGGTG
GGTGGTCGACAAGTTTATGCTCTTTCTTCTCAACACTATGCTTCATATGACCCAGCCTTCTTTATT
CATCACTCCTTTGTTGACAAAATATGGGCAGTCTGGCAAGCTCTGCAAAAGAAGAGAAAGCGTCCC
TATCATAAAGCGGATTGTGCTCTTAACATGATGACCAAACCAATGCGACCATTGTCACACGATTTTC
AATCACAATGGATTTCACAAAATGCACGCAGTCCCCAACACTCTATTTGACTTTCAGGACCTTTTC
TACACGTATGACAACTTAGAAATTGCTGGCATGAATGTTAATCAGTTGGAAGCGGAAATCAACCGG
CGAAAAGCCAAACAAGAGTCTTTGCCGGGTTCTTCTACATGGCATTGGAAGATCAGCTGATGTA
CGATTTTGGATTGTGAAGACAGCTGACGACTGCCACGCATCTGGCATGATCTTTATCTTAGGAGGT
TCTAAAGAGATGCACTGGGCCTATGACAGGAACCTTTAAATACGACATCACCCAAGCTTTGAAGGCT
CAGTCCATACACCCTGAAGATGTGTTTGACACTGATGCTCCTTTCTTCATTAAAGTGGAGGTCCAT
GGTGTAACAAGACTGCTCTCCCATCTTCAGCTATCCCAGCACCTACTATAATCTACTCAGCTGGT
GAAG

INTRON 2F-2/2G (SEQ ID NO:152)

GTGAGAGAAACTATAATAGTGTATGTCGGCAAAAAATGTGCTCATATCATGACTCTGTTGGCCGGT
GGTTGCTCTCCTCTCCTCCTCCACCACCACCGGTACCTCCACCTGTCAGGGCATCAATGTACCATG
AAAATGTCTACAATACTAGGCCTCCTGTAGAAGCACGTAAGATTTACATGGCCGGTTTGTAACTAG
TTTAAAGTGCTTCACAGTAACCAAACCAGTCTCTAAAGATTAATGTCTGTTTAAATTTAATGCC
ACATTTTCAACTGACATATTCTTGCAATTAAGTACAAATGAAGTAGTATAAATTATCCACAAATAG
CGTGATGCACCACAAATATAAACCGAGTGCTTTTTTGGCATTCCCCACTTGTTCTGGCATGATCAC
ATCATAGATCTCGTTCATGAAGATACTGTTGGATGCTTTTTCCCAATATGCCCAATCTGTTAAAT
TATTTACACGACCGCAGTGTGTACTTTCATCACTCAGATCTTTACAATGTGTTTGTAACGTTTACA
ATTAGCGTTATGATTGAAATATTACCCCTGCTACGTTAAATCACATTCACCTCACTCATCTGATGT
ACTTTACAGGTCATACCGATGATCACGGCTCAG

DOMAIN 2G-1 (1st part of domain g)

ATCATATTGCTGGCAGTGGAGTCAGGAAAGACGTGACGTCTCTTACCGCATCTGAGATAGAGAACC
TGAGGCATGCTCTGCAAAGCGTGATGGATGATGATGGACCAATGGATTCCAGGCAATTGCTGCTT
ATCACGGAAGTCCTCCCATGTGTACATGCCTGATGGTAGAGACGTTGCATGTTGTACTCATG

INTRON 2G-1/2G-2 (SEQ ID NO:153)

GTCAGTATTCTCCAATATGTTTGACTAGTGTCTTGCTCATGTATCAACTATTTTAGGCAACGTTTT
TGATTGTTATGGTATTTTCATGATATGATTTTATTGCTACCTCTATACCCAAACAAAATGTTTTA

Fig. 10e

TCAACAATTGTTTGAGTTTTAATGCAAGAAAATTATCAGGAGTAGCGTGCAAAAATGACTGGAAGG
CATGGTGTACTTCTGTGTGTACATACAAGTGGGTAATGCCTTATTGAACTCGTAATCACTCGTTTC
AG

DOMAIN 2G-2 (2nd part of domain g)

GAATGGCATCTTCCCTCACTGGCACAGACTGTTTGTGAAACAGATGGAGGATGCACTGGCTGCGC
ATGGAGCTCACATTGGCATACTACTGGGATTGGACAAGTGCGTTTAGTCATCTGCCTGCCCTAG
TGACTGACCACGAGCACAATCCCTTCCACCAC

INTRON 2G-2/2G-3 (SEQ ID NO:154)

GTCAGTATTCTCCAATATGTTTGACTAGTGTCTTGCTCATGTATCAACTATTTTAGGCAACGTTTT
TGATTGTTATGGTATTTTCATGATATGATTTTATTGCTACCTCTATACCCAAACAAAATGTTTTA
TCAACAATTGTTTGAGTTTTAATGCAAGAAAATTATCAGGAGTAGCGTGCAAAAATGACTGGAAGG
CATGGTGTACTTCTGTGTGTACATACAAGTGGGTAATGCCTTATTGAACTCGTAATCACTCGTTTC
AG

DOMAIN 2G-3 (3rd part of domain g)

GGACATATTGCTCATCGGAATGTGGATACATCTCGATCTCCGAGAGACATGCTGTTCAATGACCCC
GAACACGGGTCAGAATCATTCTTCTATAGACAGGTTCTCTTGGCTCTAGAACAGACAGACTTCTGC
CAATTTGAAGTTCAGTTTGAAATAACACACAATGCAATCCACTCTTGGACTGGAGGACATACTCCA
TATGGAATGTCATCACTGGAATATACAGCATATGATCCACTCTTTTATCTCCACCATTCCAACACT
GATCGTATCTGGGCCATCTGGCAGGCACTCCAGAAATACAGAGGTTTTCAATACAACGCAGCTCAT
TGCGATATCCAGGTTCTGAAACAACCTCTTAAACCATTTCAGCGAGTCCAGGAATCCAAACCCAGTC
ACCAGAGCCAATTCTAGGGCAGTCGATTCATTTGATTATGAGAGACTCAATTATCAATATGACACA
CTTACCTTCCACGGACATTCTATCTCAGAACTTGATGCCATGCTTCAAGAGAGAAAGAAGGAAGAG
AGAACATTTGCAGCCTTCTGTTGCACGGATTTGGCGCCAGTGCTGATGTTTCGTTTGATGTCTGC
ACACCTGATGGTCATTGTGCCTTTGCTGGAACCTTCGCGGTACTTGTTGGGGAGCTTGAGATGCCC
TGGTCCTTTGAAAGATTGTTCCGTTACGATATCACAAAGGTTCTCAAGCAGATGAATCTTCACTAT
GATTCTGAGTTCCACTTTGAGTTGAAGATTGTTGGCACAGATGGAACAGAACTGCCATCGGATCGT
ATCAAGAGCCCTACCATTGAACACCATGGAGGAG

INTRON 2G/2H (SEQ ID NO:155)

GTATGTTTTGAGATCCACATAATCTTCTACCCTGTCTCATTTCTAATGCTCTTCAATACACAATTT
ATATAGCCTTTGAGCTTCAGATGTATTACGGACAGGCATTACAGTATACATGTAATATGGTTTTCT
GCTATTTGCAAAAATTGTGTCCTATCTCTGTTTCAGATCATCATGGCGGTGACACCTAG

DOMAIN 2H (SEQ ID NO:159)

GTCACGATCACAGTGAACGTCACGATGGATTTTTTCAGGAAGGAAGTCGGTTCCTGTCCCTGGATG
AAGCCAATGACCTTAAAAATGCACTGTACAAGCTGCAGAATGATCAGGGTCCCAATGGATATGAAT
CAATAGCCGGTTACCATGGCTATCCATTCTCTGCCCTGAACATGGTGAAGACCAGTACGCATGCT
GTGTCCACGGAATGCCTGTATTTCCACATTGGCACAGACTTCATACAATCCAGTTTGAGAGAGCTC
TCAAAGAACATGGTTCTCATTTGGGTCTGCCATACTGGGACTGGAC

Fig. 11aPrimary structure of the KLH2 proteinDOMAIN B

GLPYWDWTMPMSHLPELATSETYLDPVTGETKNNPFHHAQVAFENGVTSRNPDAKLFMKPTYGDHT
YLFDSMIYAFEQEDFCDFEVQYELTHNAIHAWVGGSEKYSMSSLHYTAFDPIFYLHHSNVDRLWAI
WQALQIRRGKSYKAHCASSQEREPLKPFASFSSPLNNEKTYHNSVPTNVYDYVGVVLHYRYDDLQFG
GMTMSELEEYIHKQTQHDRTFAGFFLSYIGTSASVDIFINREGHDKYKVGSFVVLGGSKEMKWGFD
RMYKYEITEALKTLNVAVDGFSITVEITDVDGSPPSADLIPPPAIIFDVVR

DOMAIN C

ADAKDFGHSRKIRKAVDSLTVEEQTSLRRAMADLQDDKTSGGFQQIAAFHGEPPKWCPSPAEKKFA
CCVHGMVFPWHRLTLVQGENALRKHGFTGGLPYWDWTRPMSALPHFVADPTYNDSVSSLEEDNP
WYHGHIDSVGHDTTRAVRDDLYQSPGFGHYTDIAKQVLLAFEQDDFCDFEVQFEIAHNFIHALVGG
NEPYSMSSLRYTTYDPIFFLHRSNTDRLWAIWQALQKYRGKPYNTANCAIASMRKPLQPFGLDSVI
NPDETREHSVPFRVFDYKNNFDYYESLAFNGLSIAQLDRELQRRKSHDRVFAGFLLHEIGQSAL
VKFYVCKHNVSDCDHYAGEFYILGDEAEMPWRYDRVYKYEITQQLHDLDLHVGDNFFLKYEAFDLN
GGS LGGSIFSQPSVIFEPAA

DOMAIN D

GSHQADEYREAVTSASHIRKNIRDLSERGEIESIRSAFLQIQKEGIYENIAKFHGKPGLCHEHDGHPV
ACCVHGMPTFPHWHRLYLQVENALLERGSAAVAVPYWDWTEKADSLPSLINDATYFNSRSQTFDPN
PFFRGHIAFENAVTSRDPQPELWDNKDFYENVMLALEQDNFCDFEIQLELIHNALHSRLGGRAKYS
LSSLDYTAFDPVFFLHHANVDRIWAIWQDLQRYRKPYNEADCAVNEMRKPLQPFNNPELNSDSMT
LKHNLPQDSFDYQNRFRYQYDNLQFNHFSIQKLDQTIQARKQHDRVFAFGLHNIIGTSAVVDIYIC
VEQGGEQNCKTKAGSFTILGGETEMPFHFDRLYKFDITSALHKLGVPLDGHGFDIKVDVRAVNGSH
LDQHILNEPSLLFVPGERKNIYY

DOMAIN E

DGLSQHNLRKEVSSLTLEKHFLRKALKNMQADDS PDGYQAIASFHALPPLCPSPSAHRHACCL
HGMATFPQWHRLYTVQFEDSLKRHGSIVGLPYWDWLKPQSALPDLVTQETYEHLSHKTFPNPFLK
ANIEFEGEGVTTERDVAEHLFAKGNLVYNNWFCNQALYALEQENYCDFEIQFEILHNGIHSWVGG
SKTHSIGHLHYASYDPLFYIHHSQTDRIWAIWQALQEHRLSGKEAHCALEQMKDPLKPFSGSPY
NLNKRTQEFSPKPEDTFDYHRFGYEYDSLEFVGMSVSSLHNYIKQQQEADRVFAGFLLKGFGQSASV
SFDICRPDQSCQEAGYFSVLGGSSEMPWQFDRLYKYDITKTLKDMKLRYDDTFTIKVHIKDIAGAE
LDSDLIPTPSVLLEEGK

DOMAIN F

HGINVRHVGRNRIRMELSELTERDLASLKSAMRSLQADDGVNGYQAIASFHGLPASCHDDEGHEIA
CCIHGMPVFPWHRLYTLQMDMALLSHGSAAVAPYWDWTKPISKLPDLFTSPEYYDPWRDAVVNNP
FAKGYIKSEDAYTVRDPQDILYHLQDETGTSVLLDQTLLEQTDFCDFEVQFEVVHNAIHYLVGG
RQVYALSSQHYASYDPAFFIHHSFVDKIWAVWQALQKKRKPYPHKADCALNMMTKPMRPFAHDFNH
NGFTKMHAVPNTLFDQDLFYTYDNLEIAGMNVNQLEAEINRRKSQTRVFAGFLLHGIGRSADVRF
WICKTADDCHASGMIFILGGSKEMHWAYDRNFKYDITQALKAQSIHPEDVFDTDAPFFIKVEVHGV
NKTALPSSAIPAPTIISAGE

Fig. 11bDOMAIN G

DHIAGSGVRKDVTSLTASEIENLRHALQSVMDDDGPNGFQAIAAYHGSPPMCHMPDGRDVACCTHG
MASFPHWHRLFVKQMEDALAAHGAHIGIPYWDWTSASFSLPALVTDHEHNPFHGHIAHRNVDTSR
SPRDMLFNDPEHGSESFFYRQVLLALEQTDFCQFEVQFEITHNAIHSWTGGHTPYGMSSLEYTAYD
PLFYLHHSNTDRIWAIWQALQKYRGFQYNAAHCDIQVLKQPLKPFSESRNPNPVTRANSRAVDSFD
YERLNYQYDTLTFHGHISISELDAMLQERKKEERTFAAFLLHGFGASADVSDVCTPDGHCAFAGTF
AVLGGELEMPWSFERLFRYDITKVLKQMNLYDSEFHFELKIVGTDGTELPDRIKSPTIEHHGG

DOMAIN H (SEQ ID NO:158)

GHDHSEKHDGFFRKEVGSLSLDEANDLKNALYKLQNDQGPNGYESIAGYHGYPPFLCPEHGEDQYAC
CVHGMPVFPWHRLHTIQFERALKEHGSHLGLPYWDW